A Survey of Large Language Models and Machine Learning for Risk Prediction and Management in Chronic Disease Patients

www.surveyx.cn

Abstract

The integration of Large Language Models (LLMs) and machine learning technologies is revolutionizing healthcare by enhancing risk prediction and management for chronic disease patients. This survey paper explores the transformative potential of these technologies, emphasizing their role in improving clinical decision-making, personalized care, and patient outcomes. By leveraging advanced computational methods, healthcare providers can enhance predictive accuracy and develop tailored interventions aligned with individual patient needs. The integration of LLMs into healthcare systems also supports improved communication between providers and patients, facilitating informed decision-making. Despite significant advancements, challenges such as multilingual capabilities, bias, fairness, and ethical considerations persist, necessitating ongoing research and development. The survey highlights the importance of addressing these issues to ensure equitable access to AI-driven healthcare solutions. Furthermore, it underscores the need for continued innovation and collaboration among researchers, healthcare providers, and policymakers to fully realize the potential of LLMs and machine learning in transforming healthcare delivery. Overall, this paper provides a comprehensive overview of current advancements, challenges, and future directions in the application of AI technologies in healthcare, contributing to the advancement of personalized and evidence-based interventions.

1 Introduction

1.1 Importance of LLMs and Machine Learning in Healthcare

The integration of Large Language Models (LLMs) and machine learning is revolutionizing healthcare by enhancing data processing, patient care, and clinical decision-making. LLMs facilitate automated coding without labeled data, crucial for ensuring data interoperability across studies [1]. Although the application of LLMs in medicine is emergent, their potential to optimize healthcare workflows is increasingly recognized [2].

Since the introduction of ChatGPT, interest in deploying LLMs, particularly within the open-source community, has surged [3]. These models significantly improve productivity and accuracy in data science workflows, essential for healthcare [4]. LLMs' ability to generate weakly-labeled data lessens reliance on expert annotations, transforming data utilization in healthcare [5]. The emergence of Multimodal Large Language Models (MLLMs) is expected to further impact medical practice by enhancing patient interaction and data interpretation [6].

Machine learning complements LLMs by providing robust clinical decision support systems vital for managing chronic diseases. Advanced feature selection and deep learning models are reshaping healthcare, notably in predicting mortality for ischemic stroke patients [7]. Additionally, LLMs and machine learning improve the accuracy of lung cancer predictions through innovative techniques like

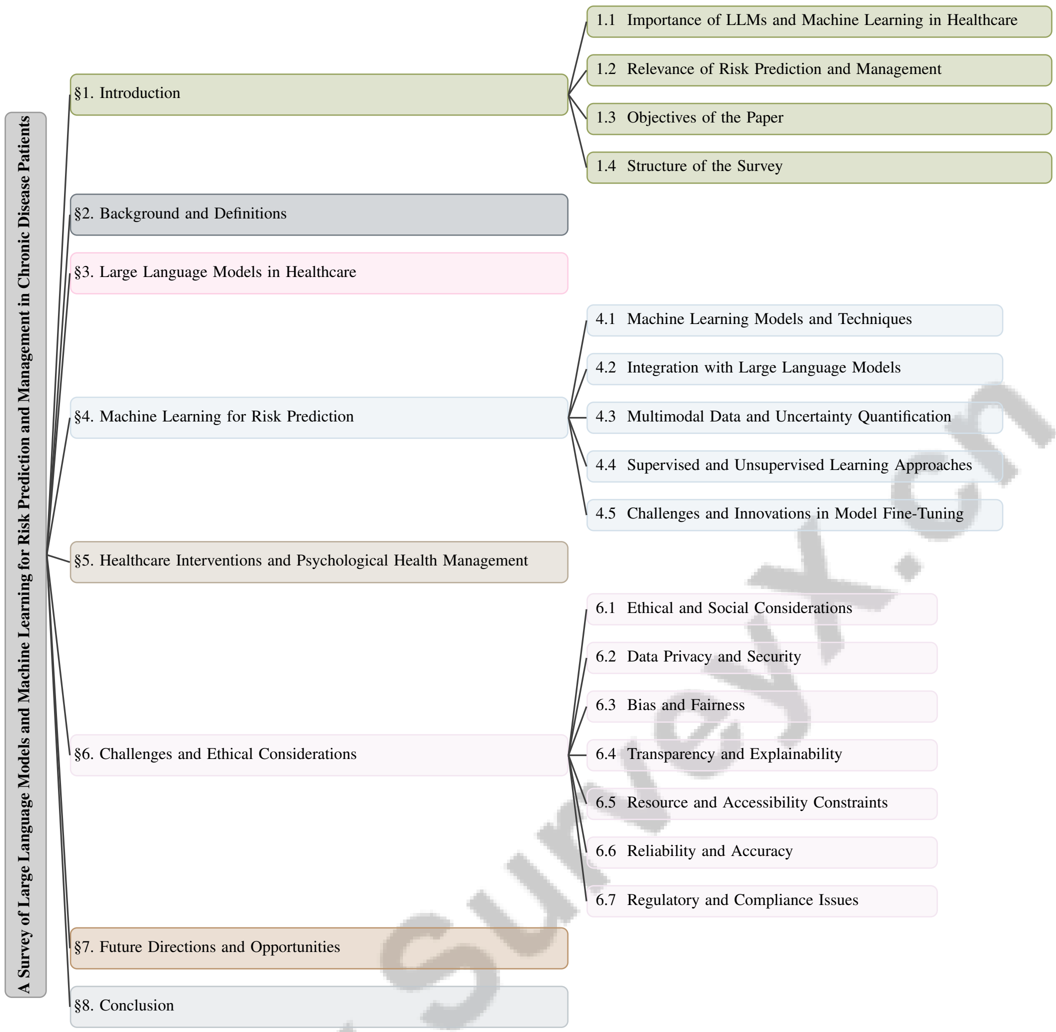


Figure 1: chapter structure

Tensor-Network Machine Learning (TN-ML) [8]. In coronary artery disease, machine learning-based diagnostic systems offer non-invasive alternatives to traditional methods like angiography [9].

Benchmarks assessing fine-tuned LLMs for generating personalized clinical impressions in PET reports advance automated medical reporting [10]. AI-driven recommender systems aid individuals in making informed health insurance decisions [11]. Integrating physiological data with LLMs may enhance empathy and interpretation of mental states, addressing current limitations in understanding psychological health [12].

Despite their potential, LLMs and machine learning face challenges such as optimizing performance on benchmarks like MMLU [13]. Innovative solutions are needed to overcome these barriers, paving the way for personalized, evidence-based healthcare. Utilizing knowledge graphs to mitigate hallucinations in LLMs enhances their reliability, critical in medical applications [14]. Furthermore, TensorFlow's robust platform is set to significantly influence machine learning research and production in healthcare [15].

1.2 Relevance of Risk Prediction and Management

Risk prediction and management are essential for optimizing healthcare outcomes in chronic disease patients, enabling early identification and intervention for potential health complications. The integration of advanced computational technologies, including LLMs and machine learning, significantly

AI-generated, for reference only.

enhances the accuracy of these predictive models. Tools like ChatGPT demonstrate their potential to improve decision-making processes in healthcare data science [4].

Effective implementation of these technologies relies on high-quality data access, a persistent challenge in healthcare [16]. This is particularly relevant for conditions like Coronary Heart Disease (CHD), where traditional diagnostic methods can be invasive or costly, highlighting the necessity for improved predictive accuracy to prevent severe health outcomes [17]. Similarly, addressing late diagnosis in Alzheimer's disease is critical for developing effective risk prediction strategies [18].

The issue of hallucinations in LLMs, where models generate plausible yet incorrect outputs, poses a significant challenge that undermines trust in healthcare applications [14]. Addressing these limitations is vital for ensuring the safety and efficacy of AI in healthcare. The integration of diverse data types through MLLMs is crucial for enhancing clinical decision-making and patient care [6].

In mental healthcare, the difficulties patients face in articulating their thoughts underscore the need for supportive tools that facilitate effective journaling practices [19]. Personalized exercise plans tailored to individual lifestyles are also essential for improving adherence and health outcomes [20]. The proposed Healthcare Copilot framework, which integrates Dialogue, Memory, and Processing components, exemplifies how AI can enhance patient-LLM interactions, supporting effective risk management and personalized care [21].

Moreover, accurately identifying causal relations from observational data is vital for effective risk prediction and management [22]. Benchmarks assessing LLM performance in clinical text summarization tasks further highlight their effectiveness compared to medical experts [23]. However, resource consumption and data scarcity remain significant barriers to the widespread adoption of MLLMs, limiting their accessibility and applicability in healthcare [24].

Inefficiencies in current dental diagnostic and treatment planning, hindered by the inability to process diverse data sources effectively, underscore the broader applicability of these technologies across healthcare domains [25]. Addressing these challenges is crucial for advancing the field and enhancing outcomes for chronic disease patients.

1.3 Objectives of the Paper

This survey paper aims to thoroughly explore the integration of LLMs and machine learning in healthcare, particularly their role in risk prediction and management for chronic disease patients. By examining the transformative potential of LLMs like ChatGPT, the paper seeks to highlight their impact on medical practices while addressing deployment challenges including generalization, evaluation metrics, and hidden costs. Additionally, the survey aims to develop a transparent framework that enables policymakers to access scientific evidence, facilitating informed healthcare decision-making [26].

The investigation will include the integration of multimodal data for chronic disease risk prediction, as proposed by novel frameworks [27], and explore lifelong learning for LLMs to prevent catastrophic forgetting, enhancing their applicability in dynamic healthcare environments [28]. Furthermore, by translating expert intuition into quantifiable metrics, the survey aims to incorporate these insights into predictive models, improving the accuracy and reliability of healthcare interventions [29].

The exploration of personalized cognitive training through AI chatbots, exemplified by the ReMe framework, will be discussed, showcasing its potential to enhance episodic memory and cognitive health [30]. The paper will also address limitations in current diagnostic methods for diseases like Alzheimer's, proposing blood-based biomarkers and computational modeling techniques to improve early diagnosis and patient outcomes [18].

Moreover, the survey will introduce the Healthcare Copilot, designed for medical consultation to address the limitations of existing medical LLMs in real-world scenarios [21]. The refinement of LLM responses through a structured framework that integrates verified medical knowledge and employs an actor-critic prompting protocol will also be discussed to enhance response accuracy [31]. Lastly, the survey will cover the deployment of LLMs on available GPU resources, assisting organizations in evaluating which models can be effectively utilized given their hardware constraints [3].

AI-generated, for reference only.

1.4 Structure of the Survey

The survey is structured to provide a comprehensive exploration of LLMs and machine learning integration in healthcare, focusing on risk prediction and management for chronic disease patients. It begins with an that emphasizes the transformative role of LLMs and machine learning technologies in healthcare, detailing their evolution from text-based systems to multimodal platforms. The subsequent section discusses the , covering applications in clinical decision support, medical imaging, and patient engagement, alongside related challenges and ethical considerations [6, 32, 33].

The survey then addresses the and clearly outlines the . The section provides foundational concepts, defining key terms and their significance in healthcare, including detailed explanations of LLMs, machine learning, risk prediction, and healthcare intervention. Following this, the specific roles of these technologies are explored, with discussing their capabilities and applications, and examining various models and techniques integrated with LLMs.

Further exploration in the survey includes , highlighting AI-driven communication, personalized interventions, and the impact on mental health. The section addresses the ethical, social, and technical challenges associated with AI in healthcare.

The paper concludes with , exploring advancements in AI capabilities and innovative deployment strategies. The final section, , summarizes the key findings and insights from the survey, emphasizing the significance of integrating LLMs and machine learning for improved healthcare outcomes. This structure ensures a logical flow of information, facilitating a thorough understanding of the topics discussed, akin to the structured approach seen in frameworks like the ReMe framework for cognitive health [30].The following sections are organized as shown in Figure 1.

2 Background and Definitions

2.1 Large Language Models (LLMs)

Large Language Models (LLMs) are reshaping artificial intelligence in healthcare by autonomously generating coherent text, enhancing clinical decision-making, and improving patient care through the analysis of vast healthcare data. Their integration with machine learning models, like Gaussian Processes and deep learning, addresses irregular sampling in electronic health records (EHRs), optimizing patient monitoring and treatment planning [34]. LLMs facilitate causal discovery, improving predictive accuracy by identifying causal relations from training data [22]. Their effectiveness is noted in specialized fields such as ophthalmology, where they provide detailed responses to complex queries [35].

In psychological health, LLMs show promise in therapy applications, such as delivering CBT-based therapy comparable to human counselors [36]. They categorize research and data sources by application areas, offering structured overviews essential for data science and machine learning [16]. Multimodal LLMs (MLLMs) expand capabilities by integrating various modalities, enhancing healthcare applications [6]. However, challenges like hallucinations in Large Vision Language Models (LVLMs) can mislead clinical diagnoses, requiring solutions to maximize LLM potential [37].

LLMs face difficulties in optimizing performance across diverse subjects, crucial for reliability in medical contexts [38]. Ongoing development, including optimization algorithms, is essential to enhance efficiency and performance in complex environments [39]. Their application in automated dental diagnosis exemplifies advancements in medical practice through natural language processing [25].

2.2 Machine Learning

Machine learning (ML) is vital in healthcare analytics, especially for predictive modeling in chronic diseases. By processing complex datasets, such as EHRs, ML identifies patterns and predicts health outcomes, notably in forecasting recovery trajectories post-stroke using neuroimaging and clinical data [40]. ML techniques, categorized into supervised and unsupervised methods, excel in scenarios with labeled and unlabeled datasets, respectively [41]. In chronic disease management, ML models enhance prediction generalizability across diverse datasets, addressing causal heterogeneity [42].

This generalizability is crucial for developing robust models for accurate predictions across various patient populations.

Challenges persist in capturing textual knowledge and generalizing to unseen tasks, especially in deep neural network-based models [43]. Addressing these involves refining ML models to support complex decision-making processes. The LEAP model exemplifies ML's application in risk prediction, optimizing intervention strategies for tuberculosis patients [44]. Operational challenges in deploying ML models include ensuring availability at the point of care, which serverless architectures can facilitate [45]. The computational burdens of fine-tuning models remain significant, necessitating solutions to maximize ML's potential in healthcare [46]. Frameworks like TensorFlow have facilitated scaling ML models across diverse hardware platforms [15].

In coronary artery disease (CAD), ML predicts disease risk over a decade, demonstrating utility in smart healthcare monitoring and advancing predictive modeling research [47].

2.3 Risk Prediction and Chronic Disease Management

Predicting health risks and managing chronic diseases are crucial for improving patient outcomes and optimizing healthcare delivery. Integrating LLMs and ML into predictive models identifies potential health risks and facilitates early interventions, addressing complexities in chronic disease management [9]. Researchers face challenges related to data application specificity, extraction complexities, and licensing [16]. Nevertheless, ML algorithms advance chronic disease management by enabling personalized treatment plans, emphasizing AI's transformative healthcare impact [25].

In psychological health, LLMs' limitations in discerning mental states hinder applications requiring emotional understanding [12]. The challenge of providing factually supported answers in specialized domains necessitates robust evidence [35]. The Med-HallMark benchmark addresses hallucinations in medical tasks, crucial for risk prediction [37]. Key challenges include data limitations, technical hurdles, and ethical considerations, like biases in training data [6]. Addressing these is essential for AI's safety and efficacy in healthcare.

Integrating LLMs and ML in risk prediction and chronic disease management holds the potential to transform healthcare practices and enhance patient outcomes. These technologies improve prediction precision and reliability, enabling effective chronic disease management. This supports personalized healthcare interventions tailored to individual characteristics, enhancing evidence-based practices through insights from data science [33, 48].

2.4 Healthcare Intervention and Psychological Health Management

AI is transformative in healthcare interventions and psychological health management, providing solutions that enhance communication and patient care. AI bridges communication gaps, particularly for cancer patients during recovery, through platforms supporting personalized care [49]. Research highlights AI's role in health communication, emphasizing message effectiveness, source credibility, and audience perception. Trustworthiness is critical for patient engagement, highlighting AI's effectiveness in enhancing communication [50]. AI empowers patients to make informed decisions, supporting better outcomes.

In mental health, AI-driven techniques transform diagnostics with explainable AI models, enhancing transparency in diagnosing and managing disorders [51]. Transparency is crucial where understanding AI-generated recommendations impacts treatment adherence. AI's integration in healthcare interventions signifies a shift toward personalized, transparent solutions. Advancements like explainable AI for disorder detection via social media leverage user-generated data for analytics. LLMs enhance problem-solving therapy, addressing mental health professional shortages. Multimodal LLMs integrate diverse data types, providing comprehensive insights into patient health. These innovations enhance mental health care efficacy while emphasizing ethical considerations [52, 53, 51, 6]. Leveraging AI capabilities enhances patient care, communication, and mental well-being, leading to better health outcomes and quality of life.

2.5 AI in Healthcare

AI is pivotal in transforming healthcare services by enhancing diagnostic accuracy, treatment planning, and patient management. AI systems process vast data, identify patterns, and provide insights supporting clinical decision-making. AI's potential is evident in predictive analytics for early disease detection, personalized medicine, robotic surgery, and telemedicine. Advancements in LLMs revolutionize patient-provider communication, enabling effective interactions in complex scenarios while streamlining documentation. These technologies enhance decision-making and patient engagement across medical domains [23, 54, 6, 35, 49].

Developing AI systems aligned with healthcare values ensures safe, high-quality outputs. Research enhances LLM safety in healthcare, encouraging exploration of value alignment [55]. AI advances precision medicine through comprehensive analyses of genetic, environmental, and lifestyle data, customizing treatment plans. Advanced ML techniques, like bandit algorithms, optimize treatment decisions based on patient-specific information, while NLP tools keep clinicians informed about genetic variants [56, 49, 48]. Personalizing care improves efficacy and reduces adverse effects, enhancing outcomes. AI-driven decision support systems assist professionals in making informed decisions with evidence-based recommendations.

AI's role encompasses improving operational efficiencies, reducing costs, and enhancing patient experiences. Integrating AI, particularly LLMs, into healthcare systems enhances communication, streamlines documentation, and improves care delivery. This addresses communication barriers, alleviates documentation burdens, and facilitates effective interactions, leading to improved outcomes and quality of life, especially in cancer recovery and mental health therapy [23, 52, 21, 49, 33].

3 Large Language Models in Healthcare

The integration of Large Language Models (LLMs) in healthcare has revolutionized patient care by leveraging advanced functionalities beyond basic data processing. This section explores the capabilities of LLMs, focusing on their technological underpinnings and effectiveness in medical contexts, which are essential for understanding their applications in healthcare settings discussed in the following subsection. Figure 2 illustrates the hierarchical structure of LLMs in healthcare, detailing their capabilities, applications, and contributions to decision-making and personalized care. Specifically, the figure highlights enhancements in processing medical data, technological advancements, and model evaluations. The applications of LLMs focus on automation, patient engagement, and clinical decision support, while the decision-making section emphasizes data-driven insights, personalized medicine, and integration with clinical systems. This visual representation not only complements the text but also provides a clear framework for understanding the multifaceted roles of LLMs in enhancing patient care.

3.1 Capabilities of Large Language Models

LLMs significantly enhance healthcare by processing complex medical data, thereby improving clinical decision-making and patient care. Domain-specific clinical models consistently outperform general-purpose models due to their tailored understanding of medical terminology and context [57]. The PaLM-2 model exemplifies advancements in LLM capabilities through selective fine-tuning, enhancing accuracy and contextual relevance [58]. Similarly, MedKP ensures reliable medical dialogue by integrating knowledge pathways [59].

The DISC-MedLLM model advances conversational healthcare services by generating accurate medical responses, crucial for maintaining communication integrity [60]. The Self-BioRAG model enhances reasoning and retrieval by focusing on biomedical information [54]. In complex medical communication, models like Sporo AraSum address challenges in medical documentation, highlighting the need for linguistic and contextual customization [61].

Statistical analyses of LLM outputs, such as attention value distribution, provide insights into model explainability, vital for understanding performance in healthcare [62]. Continuous evaluation of LLMs, including GPT-4 and open-source alternatives like FLAN-T5 and Llama-2, underscores the evolution of LLM capabilities and their healthcare applications [23].

As illustrated in Figure 3, the capabilities of large language models (LLMs) in healthcare are underscored, highlighting key advancements in model development, performance evaluation, and technical innovations. This figure emphasizes the role of fine-tuning, dialogue systems, and response generation in enhancing healthcare services, while also underscoring the importance of statistical analyses and continuous evaluation in assessing model performance and applicability.

AI-generated, for reference only.

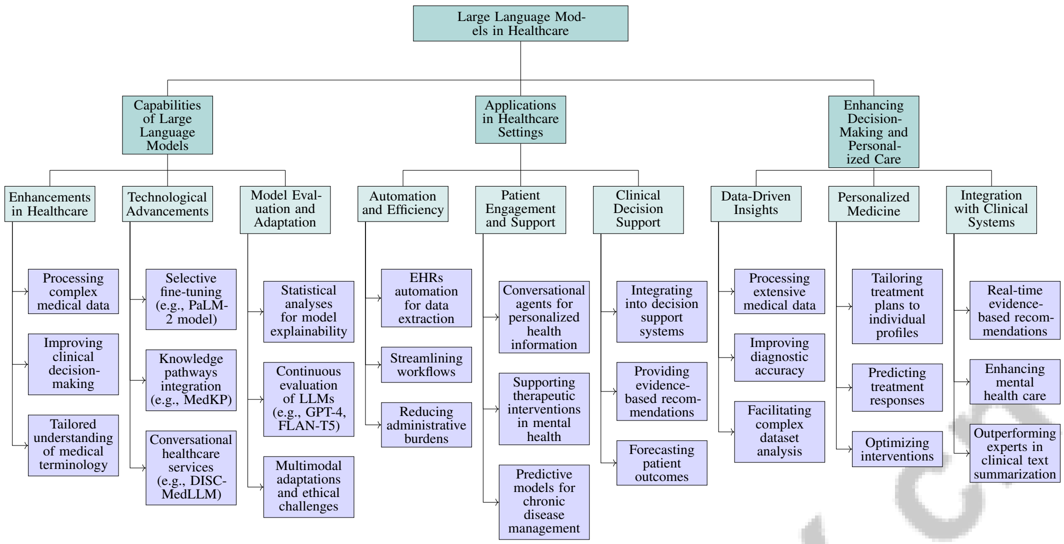


Figure 2: This figure illustrates the hierarchical structure of Large Language Models in Healthcare, detailing their capabilities, applications, and contributions to decision-making and personalized care. It highlights the enhancements in processing medical data, technological advancements, and model evaluations. The applications focus on automation, patient engagement, and clinical decision support, while the decision-making section emphasizes data-driven insights, personalized medicine, and integration with clinical systems.

Technical advancements in LLMs are driven by innovations in training, fine-tuning, and multimodal adaptations. Adapted LLMs can outperform medical experts in clinical text summarization, suggesting their potential to ease documentation burdens and enhance patient care. The evolution into Multimodal LLMs expands applications across decision support, imaging, and patient engagement, while highlighting data and ethical challenges. Not all medical adaptations enhance performance over base models, necessitating careful evaluation in medical contexts [63, 23, 6]. These advancements are crucial for improving LLMs' accuracy, reliability, and applicability in healthcare, ultimately leading to better patient outcomes and quality.

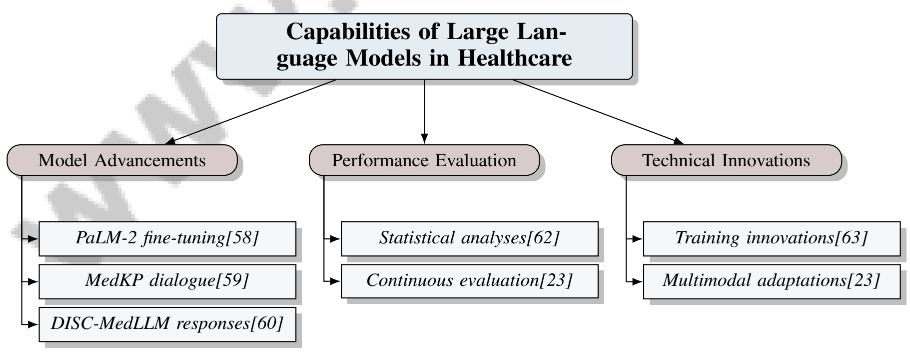


Figure 3: This figure illustrates the capabilities of large language models (LLMs) in healthcare, highlighting key advancements in model development, performance evaluation, and technical innovations. It emphasizes the role of fine-tuning, dialogue systems, and response generation in enhancing healthcare services, and underscores the importance of statistical analyses and continuous evaluation in assessing model performance and applicability.

3.2 Applications in Healthcare Settings

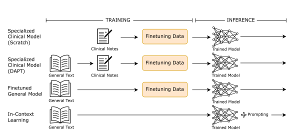
LLMs are increasingly applied in healthcare to enhance patient outcomes through various innovative methods. In electronic health records (EHRs), LLMs automate the extraction of relevant patient information, streamlining workflows and reducing administrative burdens [34]. This automation ensures critical patient data is available for clinical decision-making.

LLMs develop conversational agents that interact with patients, providing personalized health information. These agents leverage LLMs' natural language processing capabilities to improve patient engagement and adherence to treatment plans [59]. In mental health, LLMs support therapeutic interventions by generating personalized therapy content, advantageous in cognitive behavioral therapy (CBT) [36].

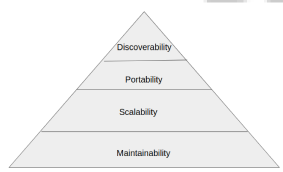
LLMs integrate into clinical decision support systems, aiding healthcare professionals in diagnosing and managing complex conditions. By analyzing medical literature and patient data, LLMs offer evidence-based recommendations for accurate diagnoses and treatment plans [57]. This integration supports personalized medicine by tailoring treatments to individual profiles.

LLMs also develop predictive models that forecast patient outcomes, particularly in chronic disease management, facilitating early intervention strategies [9]. By harnessing predictive capabilities, healthcare providers can proactively manage care, reducing complications and improving long-term outcomes.

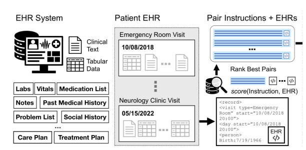
The deployment of LLMs exemplifies their transformative potential in enhancing patient care and outcomes. LLMs improve healthcare delivery through EHR management, mental health support, and clinical decision-making. In multimodal form, they integrate diverse data types for a comprehensive understanding of patient health. LLMs outperform medical experts in summarizing clinical texts, alleviating documentation burdens and enabling a greater focus on patient care. They demonstrate superior accuracy in answering health-related questions compared to traditional search engines, emphasizing their potential for timely, accurate, and personalized care. However, challenges related to data limitations, technical hurdles, and ethical considerations persist as LLMs evolve and integrate into practice [23, 6, 53, 64, 33].



(a) Training and Inference of Clinical Models[57]



(b) The image represents a pyramid with four layers, each representing a different aspect of software engineering.[65]



(c) Pairing Patient EHRs for Improved Clinical Decision-Making[66]

Figure 4: Examples of Applications in Healthcare Settings

As illustrated in Figure 4, LLM integration in healthcare is exemplified through several innovative applications. One involves training and inference of clinical models, essential for developing specialized models via domain-adaptive pretraining (DAPT). These models are fine-tuned using clinical notes to enhance accuracy in medical contexts. Another features a pyramid structure outlining key software engineering aspects crucial for deploying these models: discoverability, portability, scalability, and maintainability. This structured approach ensures models are robust and adaptable. Additionally, pairing patient EHRs with specific instructions highlights LLMs' role in improving decision-making. By aligning medical histories with tailored guidance, providers can make informed decisions, enhancing outcomes. These examples illustrate LLMs' promising applications in revolutionizing healthcare delivery and management [57, 65, 66].

3.3 Enhancing Decision-Making and Personalized Care

LLMs advance decision-making and personalized care by processing extensive medical data. They equip healthcare professionals with evidence-based insights, improving diagnostic accuracy and treatment outcomes. Integrating LLMs into workflows facilitates complex dataset analysis, such as EHRs, enabling pattern identification for patient care strategies [34].

LLMs facilitate personalized medicine by tailoring treatment plans to individual profiles. By analyzing patient data, including genetic and lifestyle factors, LLMs predict treatment responses, optimizing interventions and minimizing adverse effects [57]. This approach enhances outcomes and satisfaction by aligning treatments with individual needs.

In mental health, LLMs support personalized care by generating tailored therapeutic interventions based on patient interactions, enhancing treatment effectiveness [36]. Offering tailored feedback and tracking progress, LLMs facilitate more effective mental health care.

LLMs enhance decision-making by integrating with clinical support systems, providing real-time evidence-based recommendations and insights. This integration aids in accurate diagnoses and treatment plans, improving care and outcomes [59]. LLMs' ability to analyze complex data ensures informed decisions grounded in scientific evidence.

LLMs' contribution to decision-making and personalized care is transformative, enabling informed, accurate, and individualized care. By leveraging LLMs' capabilities, providers can enhance care quality and outcomes. LLMs excel in tasks like clinical text summarization, outperforming experts by efficiently analyzing EHRs. This alleviates documentation burdens, allowing focus on direct care and customized treatments. Multimodal LLMs, processing diverse data types, enhance decision-making and engagement. As LLMs evolve, their role in healthcare becomes vital, necessitating research to address challenges and ethical considerations [23, 67, 6, 35, 33].

4 Machine Learning for Risk Prediction

Category	Feature	Method
Machine Learning Models and Techniques	Human-Centric Optimization Explanation and Insight Methods Quantum and Advanced Encoding	iACO[68] ITP-XML[69] TN-ML[8]
Integration with Large Language Models	Task-Specific Pre-training Data Integration Strategies	SMP-BERT[70] MMLLMAIS[25]
Multimodal Data and Uncertainty Quantification	Multimodal Data Fusion	CTF[71], m2d2[72], LLM-AD[73]
Supervised and Unsupervised Learning Approaches	Integrated Learning Strategies	HRNN[40]
Challenges and Innovations in Model Fine-Tuning	Data Utilization Strategies Test Feature Adjustment Task-Specific Enhancement Resource Optimization	TEE4EHR[34] CA[74] TSKD[13] Llama-SFTn-WS-BERTn[5]

Table 1: This table provides a comprehensive overview of various machine learning models and techniques employed in healthcare risk prediction. It categorizes these methods into key areas such as machine learning models and techniques, integration with large language models, multimodal data and uncertainty quantification, supervised and unsupervised learning approaches, and challenges and innovations in model fine-tuning. Each category is further detailed with specific features and methodologies, highlighting the diverse applications and innovations shaping the field.

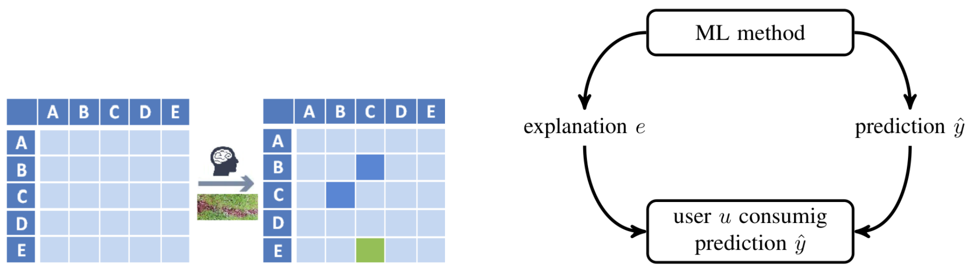
Machine learning (ML) techniques are pivotal in enhancing risk prediction within healthcare analytics. Table 3 offers a detailed classification of machine learning models and techniques crucial for advancing risk prediction in healthcare analytics. This section examines diverse ML models and methodologies that identify and forecast health risks, facilitating timely interventions and improved patient outcomes. By evaluating traditional algorithms alongside advanced methodologies, we gain insight into their roles in the evolving landscape of healthcare risk assessment. The following subsection will detail specific ML models and techniques currently shaping risk prediction.

4.1 Machine Learning Models and Techniques

ML models are crucial for advancing risk prediction in healthcare, utilizing sophisticated methodologies to analyze complex datasets. Traditional algorithms, such as Random Forest, Logistic Regression, and Support Vector Machines, have demonstrated efficacy in clinical settings for disease prediction, including coronary heart disease [17]. These models manage large datasets and complex patterns while offering interpretability [47]. Advanced architectures like Convolutional Neural Networks (CNNs) and Long Short-Term Memory networks (LSTMs) enhance capabilities by analyzing sequential and spatial data, critical in healthcare applications [15]. Integrating domain knowledge with ML models further enhances reasoning capabilities, as seen in the LLM-Powered Weak Supervision method, which uses Large Language Models (LLMs) to generate weak labels from unannotated clinical notes, training a BERT model [5].

Specific disease risk prediction models, such as the Tensor-Network Machine Learning (TN-ML) model, have shown high accuracy in predicting lung cancer stages [8]. The TEE4EHR model, a transformer-based approach, encodes event sequences in electronic health records (EHRs) and addresses irregular sampling through a point process framework [34]. Benchmark evaluations of models like Random Forest, Multilayer Perceptron, and AdaBoost highlight their diverse applications in intelligent decision support systems [9]. These methodologies enhance predictive accuracy and support personalized care strategies, though challenges remain, such as the high computational demands of LLMs and the need for extensive training datasets [39].

Innovative methodologies, like LLMs and retrieval-augmented generation (RAG), improve predictive accuracy, leading to better patient outcomes and more effective personalized care strategies. Contextual bandit algorithms in precision medicine allow tailored treatment decisions based on individual characteristics, refining health risk assessments and enhancing clinical decision-making [23, 48, 10, 35, 33].



(a) Brainstorming Session[68]

(b) ML method[69]

Figure 5: Examples of Machine Learning Models and Techniques

As illustrated in Figure 5, various models and techniques in machine learning for risk prediction enhance decision-making processes. The "Brainstorming Session" image visualizes a conceptual framework where potential ideas are represented in a grid format, emphasizing the iterative development of ML models. The "ML method" image presents a flowchart depicting the interaction between a machine learning method and its user, highlighting the cyclical nature of generating predictions and interpreting them through explanations, underscoring the importance of transparency and user feedback in refining predictive models [68, 69].

4.2 Integration with Large Language Models

Integrating ML models with Large Language Models (LLMs) transforms healthcare by enhancing the processing and interpretation of complex datasets for improved diagnostic accuracy and patient care. This integration leverages ML and LLM strengths, facilitating comprehensive risk prediction and management strategies. The MultiModal LLM AI System (MMLLMAIS) exemplifies this by utilizing a large language model to analyze multi-source data for automated dental diagnosis and treatment planning [25].

Innovative techniques such as adapter tuning tailor ML models for medical images, ensuring seamless collaboration with LLMs and enhancing healthcare solutions. Frameworks like MAKEN utilize structured pre-training tasks to discern logical relationships within medical reports, improving information extraction and diagnostic processes [70]. In medical coding, integrating ML models with LLMs, as demonstrated by the LLM-codex method, enhances ICD code prediction accuracy, streamlining the coding process and reducing errors. Specialized models like SMP-BERT, pre-trained on structured tasks, effectively extract and synthesize medical information [70].

Meta-learning approaches, such as Bayesian hierarchical modeling, illustrate the potential of integrating ML with LLMs to enhance generalizability and predictive accuracy across related tasks. This is complemented by models utilizing convolutional neural networks trained on hybrid images, merging neuroimaging and tabular data to predict recovery trajectories following stroke [40]. The Random Forest model, achieving an accuracy of 84

4.3 Multimodal Data and Uncertainty Quantification

Integrating multimodal data in healthcare analytics is essential for enhancing predictive model accuracy and robustness, particularly in risk prediction and chronic disease management. Multimodal approaches leverage diverse data sources, such as electronic health records (EHRs), imaging, and narrative text, to capture comprehensive patient profiles and improve diagnostic accuracy. Constrained tensor factorization methods effectively capture complex interactions in EHR data, ensuring that derived phenotypes are interpretable and predictive of outcomes [71].

Utilizing informative priors, such as the multimodal data-driven (m2d2) prior, significantly enhances the reliability and predictive performance of multimodal clinical data classification [72]. This ensures that predictive models are accurate and capable of generalizing across different patient populations and clinical settings. The Random Forest model, achieving a classification accuracy of 95.7

Integrating LLMs with anomaly detection techniques further demonstrates the utility of multimodal data, as LLM-AD adapts to specific patterns of anomalies in computational workflows [73]. This integration enhances the predictive accuracy of healthcare models by identifying subtle patterns indicative of underlying health issues. In analyzing verbal autopsy reports, the fusion of binary features and narrative text creates a comprehensive dataset for assessing causes of death, such as uncontrolled hyperglycemia [75]. This approach underscores the significance of incorporating diverse data types to improve the interpretability and predictive power of healthcare models.

Uncertainty quantification is vital in predictive modeling within healthcare, enhancing the reliability of model predictions and providing insights into confidence levels. This process integrates expert domain knowledge through advanced methodologies, such as LLMs, and employs techniques like Bayesian inference to systematically address variability in multimodal clinical data. By improving predictive model robustness, uncertainty quantification supports better decision-making and risk assessment in patient care, ensuring that healthcare analytics are informed and actionable [76, 29, 72, 48, 77]. Leveraging multimodal data enables healthcare providers to better assess prediction uncertainties, facilitating informed decision-making and personalized care strategies.

4.4 Supervised and Unsupervised Learning Approaches

Supervised and unsupervised learning approaches are fundamental to healthcare risk prediction, each offering unique advantages in analyzing complex datasets to forecast patient outcomes and manage chronic diseases. Supervised learning relies on labeled datasets to train models for predicting specific health outcomes, proving particularly effective when historical data is available. This approach has been instrumental in developing predictive models for chronic diseases, utilizing algorithms like Random Forest and Support Vector Machines to identify risk factors and predict disease progression [17].

Conversely, unsupervised learning excels at uncovering hidden patterns within healthcare data that may not be immediately apparent. This approach is valuable in exploratory data analysis, where the aim is to identify clusters or associations without predefined labels. Techniques such as clustering and dimensionality reduction reveal insights into patient subgroups and disease phenotypes, facilitating personalized treatment plans. The capacity of unsupervised learning to detect anomalies and uncover novel patterns is crucial for enhancing healthcare analytics and improving patient outcomes [47].

The integration of supervised and unsupervised learning approaches is exemplified in predicting recovery trajectories following stroke. Here, supervised learning models utilize labeled neuroimaging and clinical data to forecast language deficits, while unsupervised techniques identify underlying patterns in patient recovery [40]. This combination enhances predictive model accuracy and robustness, ensuring timely and effective healthcare interventions.

Integrating supervised and unsupervised learning methodologies is vital for enhancing healthcare risk prediction, as supervised learning excels in outcome classification and prediction, while unsupervised learning uncovers hidden patterns and relationships within complex datasets, leading to more accurate and comprehensive risk assessments [76, 78, 56, 77, 70]. By leveraging these methodologies, healthcare providers can enhance predictive accuracy, improve patient outcomes, and support personalized care strategies, advancing evidence-based healthcare interventions.

4.5 Challenges and Innovations in Model Fine-Tuning

Method Name	Computational Challenges	Bias and Reliability	Innovative Adaptations
TSKD[13]	Extensive Computational Resources	-	Two-stage Knowledge Distillation
CA[74]	-	Selection Biases	Causality-aware Approach
Llama-SFTn-WS-BERTn[5]	Computational Burden	Potential Biases	Weak Supervision
TEE4EHR[34]	Hyperparameter Tuning Required	Clinical Prediction Tasks	Point Process Framework

Table 2: Summary of computational challenges, biases, and innovative adaptations in various model fine-tuning methods for healthcare applications. The table highlights the specific challenges each method faces, such as computational resource demands and potential biases, alongside the innovative strategies employed to enhance model performance and reliability.

Fine-tuning ML models for healthcare applications presents several challenges, alongside innovations aimed at enhancing model performance and applicability. A primary challenge is the extensive computational resources required for training, which can lead to overfitting, particularly during the refinement phase, necessitating careful hyperparameter tuning to ensure model robustness [13]. Existing methods often focus solely on deconfounding training data, neglecting the need to deconfound test set features for achieving stable predictions, posing a significant challenge for reliable model outputs [74]. Table 2 provides a comprehensive overview of the computational challenges, biases, and innovative adaptations associated with different model fine-tuning methods in healthcare applications.

Reliance on LLMs for detection and integration tasks introduces potential biases and limitations in scalability and generalizability, as studies often depend heavily on these models [14]. Medical hallucinations present a critical challenge, as benchmarks fail to capture their complexities, relying on outdated datasets and metrics that do not accurately reflect the factual correctness of model outputs [37]. The MediConfusion framework highlights the struggle of medical MLLMs to differentiate between confusing image pairs, raising concerns about their reliability in healthcare applications [79].

Innovations in model fine-tuning have focused on addressing these challenges by improving evaluation and adaptation of LLM capabilities. Domain-adapted evaluation metrics, tailored for tasks like PET impression generation, allow for more accurate assessment of model performance compared to existing general metrics [10]. Additionally, leveraging LLMs for weak supervision in healthcare applications demonstrates lower computational costs compared to traditional LLM approaches while improving performance with minimal labeled data [5].

The TEE4EHR framework exemplifies an innovative approach by capturing the informative nature of missing data in EHRs, improving prediction accuracy and addressing challenges in traditional methods that rely on imputation [34]. Furthermore, using structured knowledge bases mitigates representational heterogeneity in metadata, ensuring that ML models remain robust and reliable, thus contributing to improved healthcare solutions.

To effectively tackle challenges in predictive analytics and automatic grading, innovative strategies for model fine-tuning are essential. This includes improving imbalanced dataset management and enhancing LLM interpretability. Recent advancements highlight the potential of LLMs to incorporate expert domain knowledge into quantifiable features, significantly boosting model performance in risk assessment and decision-making. Employing Parameter Efficient Fine-Tuning (PEFT) methods, such as LoRA and QLoRA, optimizes LLMs for automatic grading tasks while reducing computational demands. By focusing on these innovative approaches, we can enhance both model accuracy and clarity of outputs, leading to more effective applications across various fields [80, 81, 29, 32]. Overcoming these challenges and leveraging recent innovations can fully realize the transformative potential of ML models in healthcare delivery and patient outcomes.

5 Healthcare Interventions and Psychological Health Management

5.1 AI-Driven Health Communication

AI-driven health communication is transforming healthcare by enhancing patient-provider interactions, fostering engagement, and improving decision-making. Large Language Models (LLMs) and Multimodal Large Language Models (MLLMs) have revolutionized medical applications, facilitating

Feature	Machine Learning Models and Techniques	Integration with Large Language Models	Multimodal Data and Uncertainty Quantification
Data Handling	Complex Datasets	Complex Datasets	Multimodal Data
Model Type	Traditional And Advanced	Llm-enhanced	Multimodal
Unique Feature	Integrates Domain Knowledge	Improves Diagnostic Accuracy	Enhances Predictive Robustness

Table 3: This table provides a comprehensive comparison of various machine learning models and techniques used in healthcare risk prediction. It highlights key features such as data handling, model types, and unique attributes, including the integration of domain knowledge, enhancement of diagnostic accuracy through large language models, and the robustness of predictive capabilities with multimodal data.

clinical diagnosis and patient interaction, thereby elevating healthcare quality. These models enable real-time insights into patient health, integrating vision, audio, and language inputs for comprehensive assessments and recommendations [25].

The integration of AI technologies, such as continuous health monitoring systems, enhances communication between patients and providers by enabling ongoing assessments and timely interventions. AI-driven benchmarks provide non-invasive alternatives to traditional diagnostics, improving access to accurate diagnoses and facilitating better health information communication. For instance, benchmarks for evaluating LLMs in generating personalized medical impressions enhance clinical workflows and decision-making, boosting patient engagement [10].

AI systems like Healthcare Copilot streamline medical consultations, allowing providers to focus on patient care. Retrieval-Augmented Generation (RAG) methods have proven effective in generating high-quality feedback while reducing computational costs, enhancing scalability, accuracy, and reliability of patient education materials. Recent advancements in domain-specific applications, particularly in ophthalmology, demonstrate that integrating RAG with large language models can improve factual accuracy in responses from 20.6

In mental health management, tools utilizing LLMs promote deeper patient engagement through conversational prompts, enhancing emotional expression and insight. This method aids patients in articulating their thoughts and emotions, essential for effective therapeutic interventions, as studies indicate that improved communication fosters therapeutic alliance and empathy in both human and AI-assisted therapy settings [52, 19, 36].

AI-driven health communication marks a significant advancement in healthcare, fostering effective provider-patient interactions, enhancing diagnostic accuracy, and enabling personalized care. Recent LLM advancements, such as Med-PaLM and BioBERT, have outperformed previous benchmarks in medical question answering and text summarization, improving information exchange quality in clinical settings. These models address communication barriers faced by patients, particularly in complex scenarios like post-cancer treatment recovery, while streamlining clinical documentation processes, allowing professionals to focus more on direct patient care. The integration of AI technologies in health communication is poised to significantly enhance patient outcomes and healthcare delivery efficiency [82, 23, 49, 33].

5.2 Personalized Healthcare Interventions

AI, particularly through Large Language Models (LLMs), significantly advances personalized healthcare interventions by analyzing extensive datasets to generate tailored solutions for individual patient needs. For example, AI-driven approaches for children with Autism Spectrum Disorder (ASD) create customized verbal content, enhancing intervention effectiveness by addressing specific communication and social interaction challenges [83].

Technologies like PlanFitting exemplify personalized healthcare by allowing users to express lifestyle constraints and receive customized exercise recommendations, ensuring alignment with individual circumstances and health goals, thereby enhancing patient adherence and outcomes [20].

AI's transformative role in personalized healthcare interventions enables more effective, customized care solutions. By tailoring interventions to meet individual patient needs, AI enhances healthcare quality and outcomes. This patient-centered approach is particularly evident in cancer patient-provider communication, where AI bridges knowledge gaps and facilitates improved dialogue during recovery. Additionally, LLMs are being explored as therapeutic tools in psychotherapy, showcasing their potential to provide personalized problem-solving therapy, especially amid the shortage of

AI-generated, for reference only.

mental health professionals. These advancements underscore AI's transformative role in fostering a responsive and effective healthcare delivery system [52, 23, 49].

5.3 Mental Health and Psychological Management

AI integration in mental health care and psychological management signifies substantial advancements in diagnosing and treating mental disorders. Current research highlights machine learning models' potential to predict mental health outcomes, offering insights into mental disorder complexities and paving the way for effective interventions [76]. Frameworks like DKDD provide valuable insights into factors contributing to depression, enabling timely interventions and enhancing psychological health management [84].

Large Language Models (LLMs) have significantly advanced therapeutic settings by addressing the shortage of mental health professionals and providing scalable care solutions [52]. The interpretability and accuracy of models like LLMAD enable users to understand detected anomalies, crucial for informed decision-making in mental health contexts [85]. This transparency fosters trust and acceptance of AI-driven tools among healthcare providers and patients.

Integrating physiological data with LLMs enhances empathic interactions, as pilot studies validate this approach [12]. By incorporating physiological signals, AI systems can better respond to patients' emotional states, providing nuanced psychological support. Furthermore, the TN-ML approach enhances diagnostic reliability by flagging uncertain predictions for human expert review, critical in mental health contexts [8].

AI's impact on mental health extends to detecting disorders through social media, where explainable AI models provide transparency and reliability in identifying potential issues [51]. Advancements in interactive AI systems, such as IvyGPT, show significant improvements in generating medically relevant responses, indicating potential enhancements in self-service healthcare and support for medical professionals in mental health management [86].

The integration of AI in mental health care offers transformative potential, enabling accurate diagnoses, personalized interventions, and enhanced patient-provider interactions. By leveraging AI's advanced capabilities, particularly through explainable AI and LLMs, the healthcare industry can effectively address persistent mental health challenges, improving service quality and accessibility. Recent research emphasizes AI-driven analytics using social media data to enhance mental disorder detection, highlighting the importance of transparency and interpretability in AI applications for informed decision-making. Moreover, the comparative performance of AI models, such as Llama-2 and ChatGPT, against traditional machine learning approaches underscores the evolving landscape of mental health solutions, paving the way for innovative strategies to support the over 20

5.4 Integration with Existing Healthcare Systems

Integrating AI technologies with existing healthcare systems is crucial for enhancing care quality and operational efficiency. By embedding AI-driven solutions into healthcare infrastructures, institutions can significantly improve various aspects of patient care, from diagnosis to treatment and resource management. For instance, TEE4EHR demonstrates potential applications in healthcare by improving representation learning in electronic health records (EHRs) and enhancing clinical prediction tasks [34].

AI technologies like LLM-codex have been integrated into healthcare systems to improve medical coding accuracy, supporting clinical decision-making and reducing administrative burdens. Advanced frameworks replicating human reasoning in medical coding, such as Self-BioRAG and ALIGN, incorporate essential validation steps that minimize errors and enhance trust in automated outputs while facilitating seamless integration with existing systems. Self-BioRAG employs retrieval-augmented generation methods to enhance biomedical text generation and validation, while ALIGN improves automated medical coding through a structured three-step process, ensuring reliability and interoperability in clinical trial data integration [1, 23, 54, 53, 32]. These advancements illustrate AI's potential to streamline data management processes, improving healthcare delivery efficiency and accuracy.

Moreover, AI-powered frameworks reduce reliance on costly expert annotations while achieving high accuracy in detecting structured data elements, making them scalable solutions for healthcare interventions. Recent advancements in AI, particularly through LLMs, highlight their ability to enhance data management processes significantly. These models excel in clinical text summarization, outperforming medical experts in certain tasks, thereby alleviating clinicians' documentation burden. By streamlining the extraction and summarization of critical information from electronic health records, LLMs improve healthcare delivery efficiency and accuracy. The integration of structured evaluation frameworks for clinical summaries ensures reliable outputs conducive to better patient-provider communication, suggesting transformative potential for AI technologies in optimizing healthcare workflows and supporting informed decision-making in clinical settings [23, 82, 51, 53, 49].

In pharmacovigilance, AI approaches empower non-technical users to perform complex data queries, making critical information more accessible and facilitating timely healthcare interventions. This democratization of data access enables healthcare providers to make informed decisions through comprehensive data analyses, aided by technologies like LLMs that can automatically extract and interpret medical decision trees from texts, convert natural language queries into precise SQL queries for pharmacovigilance data, and analyze health opinions in online communities. Integrating diverse data sources and analytical methods enhances clinical decision-making accuracy and relevance, ultimately improving patient outcomes [87, 78, 77, 48].

The integration of self-supervised learning methods, particularly LLMs, in healthcare settings has shown significant potential for enhancing clinical decision-making, especially in resource-constrained environments. Studies indicate that adapted LLMs can outperform medical experts in clinical text summarization tasks, such as radiology reports and patient interactions, reducing documentation burdens and allowing clinicians to prioritize patient care. Frameworks like Self-BioRAG have demonstrated improvements in medical reasoning by effectively retrieving and reflecting on relevant biomedical information, while novel approaches like Text2MDT facilitate the automatic extraction of medical decision trees from texts, streamlining clinical decision support system development. These advancements highlight self-supervised learning's transformative role in improving healthcare delivery and decision-making processes [78, 54, 23]. These methods have achieved performance levels comparable to experienced radiologists in predicting patient deterioration, underscoring their potential to enhance diagnostic accuracy and patient care.

Structured knowledge bases are also critical for improving metadata quality, addressing adherence and consistency issues, and ensuring that AI models remain robust and reliable. By integrating advanced tools like LLMs into healthcare systems, organizations can significantly enhance data management quality and streamline clinical workflows, alleviating the documentation burden on clinicians. Recent studies indicate that adapted LLMs can outperform medical experts in summarizing clinical texts, such as radiology reports and patient interactions, enhancing clinical documentation accuracy and efficiency. Frameworks like Attribute Structuring improve the evaluation of these summaries, ensuring reliable information dissemination. As LLMs continue to demonstrate superior capabilities in answering health-related questions compared to traditional search engines, their implementation can facilitate more effective patient care and decision-making processes in healthcare settings [53, 23, 66, 33].

The integration of AI technologies with existing healthcare systems is transformative, enhancing patient care, improving operational efficiency, and supporting personalized healthcare interventions. Recent advancements in AI, particularly in LLMs and explainable AI, are significantly reshaping healthcare delivery by enhancing patient-provider communication and improving mental health diagnostics through social media analysis. These technologies address critical challenges, including communication barriers in post-cancer treatment recovery and the need for transparency in mental health assessments, ultimately shaping a more efficient and responsive healthcare system [51, 49].

6 Challenges and Ethical Considerations

The rapid evolution of artificial intelligence (AI), particularly with the advent of Multimodal Large Language Models (MLLMs), demands careful scrutiny of the challenges and ethical implications associated with their integration into healthcare. These models are pivotal in transforming clinical decision-making and patient engagement, yet they introduce concerns about data limitations and the necessity for explainability, especially in sensitive areas like mental health analytics. Addressing these multifaceted issues is crucial for responsible AI deployment in medical practice [51, 6]. The use of Large Language Models (LLMs) raises critical questions about reliability, bias, and their overall impact on patient care, which are essential for understanding the complexities of responsible AI use in clinical environments and guiding equitable healthcare solutions.

6.1 Ethical and Social Considerations

The integration of AI in healthcare, particularly through LLMs, presents significant ethical and social considerations. A primary concern is the potential for LLMs to generate hallucinations, undermining the reliability and accuracy of medical decision-making [37]. Tools like the Med-HallMark and MediHall Score have emerged to evaluate hallucination detection in medical LLMs, addressing ethical considerations related to reliability and clinical applicability [37].

The complexity and non-interpretability of existing machine learning models, especially deep learning approaches, raise ethical concerns regarding trust and reliability in clinical settings [47]. Studies often struggle with the interpretability of machine learning models and the assumptions required by statistical methods, which can limit their applicability [47]. The subjective nature of explainability and the absence of standardized metrics for evaluating effectiveness further complicate the ethical landscape of AI in healthcare [88].

Bias in AI outputs is another critical issue, affecting the fairness and inclusivity of AI-driven healthcare solutions. Research limitations involve biases in training data, lack of standardized evaluation metrics, and challenges in achieving interpretability and transparency in model decision-making [6]. This concern is exacerbated by the potential for AI models to produce outputs that do not represent the entire population, highlighting the necessity for comprehensive evaluation frameworks that accurately assess LLM performance across diverse linguistic and cultural contexts [35].

Privacy concerns are paramount in the ethical discourse surrounding AI in healthcare, especially regarding the handling of sensitive patient data. Future research should focus on addressing privacy and ethical concerns related to sensitive health data [12]. Ensuring that data used in AI applications is appropriately licensed and anonymized is crucial for maintaining ethical standards [25]. The use of AI in therapeutic settings emphasizes ethical implications concerning empathy and the handling of sensitive topics, underscoring the need for guidelines governing AI deployment in healthcare.

The ethical and social implications of employing AI in healthcare underscore the urgent necessity for comprehensive frameworks that ensure transparency, promote equitable access to AI technologies, and uphold rigorous ethical standards. These frameworks are essential for addressing the complexities of mental health analytics, responsible LLM deployment, and multimodal data integration in clinical practice, fostering trust and accountability among stakeholders while navigating the challenges and opportunities presented by AI in healthcare [89, 51, 6, 53, 49]. By addressing these challenges, the healthcare industry can harness the potential of AI while maintaining trust and social responsibility.

6.2 Data Privacy and Security

Data privacy and security are paramount in deploying AI-driven healthcare solutions, particularly with the integration of LLMs and MLLMs. The development of medical MLLMs faces critical challenges, including ensuring the security and freshness of healthcare data, which directly impacts the quality of outputs generated by these models [90]. The handling of sensitive patient information necessitates robust privacy protection mechanisms to prevent unauthorized access and data breaches.

A significant issue is the potential for LLMs to inadvertently memorize and disclose private data. Metrics have been developed to evaluate models' accuracy in refusing to disclose sensitive information and their likelihood of memorizing private data, emphasizing the importance of safeguarding patient confidentiality [91]. Ensuring that AI systems do not compromise patient privacy is crucial for maintaining trust and compliance with regulatory standards.

Additionally, the deployment of LLMs must consider potential biases in predictions arising from user demographics. These biases may affect the accuracy of personality inferences and other predictive outcomes, underscoring the need for evaluation and mitigation strategies to ensure fair and unbiased AI applications in healthcare [92].

Organizations deploying LLMs in healthcare must balance performance and resource constraints. The benchmark emphasizes the importance of ensuring that AI solutions remain effective and efficient without compromising data privacy and security [3]. By addressing these challenges, healthcare providers can leverage AI technologies while safeguarding patient data and maintaining ethical standards in AI-driven healthcare solutions.

6.3 Bias and Fairness

Bias and fairness are critical concerns in deploying AI models, particularly LLMs, in healthcare. These concerns stem from inherent biases in LLM training data, which can lead to skewed outcomes and perpetuate historical stereotypes, negatively impacting marginalized groups [93]. Understanding potential biases in LLMs is crucial for assessing fairness in AI models, as they influence the reliability and equity of healthcare solutions [22].

A significant challenge lies in models like ChatGPT's ability to generalize knowledge and demonstrate reasoning across diverse contexts. This limitation can result in categorical failures, affecting the model's ability to provide equitable healthcare solutions [94]. The generalization capabilities of LLMs are not yet fully understood, posing a risk of biased outputs that may not accurately reflect the diverse needs of different patient populations.

Addressing these biases requires a comprehensive approach that involves analyzing and mitigating hidden stereotypes encoded within LLMs. This process is essential to ensure AI-driven healthcare solutions do not inadvertently reinforce existing disparities or introduce new inequities in patient care. Continuous assessment and enhancement of training datasets are necessary to ensure they adequately reflect diverse populations and promote inclusivity, thereby mitigating the risk of perpetuating biases inherent in LLMs and other AI systems. Recent research emphasizes the importance of diverse data representation in improving model performance and healthcare outcomes [95, 51, 93, 96, 32].

Addressing bias and fairness in AI models utilized in healthcare is essential to ensure these technologies promote equitable access to healthcare services, enhance patient outcomes, and mitigate the risk of perpetuating societal stereotypes or disparities. By fostering transparency and interpretability in AI systems, we can better understand their internal processes and biases, ultimately leading to more ethical and effective applications in critical areas such as mental health and patient-provider communication [93, 49, 51]. By addressing these concerns, the healthcare industry can harness AI's potential while upholding ethical standards and promoting social justice.

6.4 Transparency and Explainability

Transparency and explainability are critical components in deploying AI healthcare applications, directly impacting trust, adoption, and the efficacy of AI-driven solutions. The complexity and opacity of LLMs pose significant challenges in healthcare, where understanding AI systems' decision-making processes is crucial for ensuring accurate and reliable patient care. A human-centered framework for transparency categorizes existing approaches and emphasizes the specific needs of various stakeholders in healthcare AI applications [89]. This framework highlights the importance of tailoring transparency efforts to meet the diverse requirements of patients, healthcare providers, and policymakers.

Explainability in LLMs is particularly challenging due to the sensitivity of their explanations to the randomness inherent in their training processes [62]. This randomness can lead to variations in model outputs, complicating consistent interpretation and trust in AI-generated recommendations. The need for explainability is underscored by the potential for LLMs to produce outputs that are not easily understandable by end-users, thereby hindering AI integration into clinical workflows.

A taxonomy of explainable supervised machine learning (SML) methods categorizes these into interpretable models, surrogate model fitting, and explanation generation, providing a structured approach to enhancing AI systems' transparency [88]. By employing these methods, healthcare AI applications can offer clearer insights into their decision-making processes, increasing user trust and facilitating AI technology adoption in clinical settings.

The pursuit of transparency and explainability in AI healthcare applications is essential for fostering trust, ensuring ethical use, and promoting effective AI integration into healthcare systems. By overcoming communication barriers and resource limitations identified in the post-cancer treatment recovery phase, the healthcare industry can effectively harness AI technologies, particularly LLMs, to improve patient-provider interactions, streamline clinical documentation, and uphold ethical standards, ultimately enhancing patient care quality [51, 23, 49, 6].

6.5 Resource and Accessibility Constraints

The implementation of AI solutions in healthcare is significantly hindered by resource and accessibility constraints, affecting the scalability and effectiveness of these technologies. A critical challenge is the high computational demand associated with training and deploying LLMs, necessitating substantial hardware resources and energy consumption, thereby limiting applicability in resource-constrained environments [24]. The reliance on high-quality training data exacerbates this issue, as the availability of diverse and comprehensive datasets is essential for developing robust AI models capable of generalizing across different healthcare contexts [4].

The complexity of health information needs often exceeds current AI models' capabilities, which tend to rely on binary questions that fail to capture the nuanced requirements of healthcare decision-making [33]. This limitation underscores the necessity for more sophisticated models capable of dynamic knowledge representation and comprehensive analysis, moving beyond static templates to address the diverse needs of healthcare providers [93].

Moreover, the limitations of existing benchmarks, such as ALIGN, in handling uncommon codes highlight the need for human intervention and further refinement to ensure accurate coding and data processing [1]. The reliance on single datasets for model evaluation, as seen in studies predicting coronary heart disease, raises concerns about the generalizability of findings across different populations, necessitating the inclusion of diverse datasets to enhance model robustness [17].

The demand for explainability in AI systems, driven by regulations such as the EU General Data Protection Regulation (GDPR), presents additional challenges in ensuring that AI solutions are both transparent and compliant with legal standards [88]. The comprehensive framework provided by TensorFlow facilitates the evaluation of machine learning models across diverse hardware platforms, promoting innovation and research in the field [15]. However, the resource-intensive nature of these evaluations remains a barrier to widespread adoption.

Addressing resource and accessibility constraints in healthcare is essential for unlocking the full potential of AI technologies, particularly LLMs and MLLMs. These technologies can enhance communication between patients and providers, improve clinical decision-making, and facilitate patient engagement by integrating diverse data types. To achieve scalable, efficient, and equitable patient care, it is vital to develop AI applications with a focus on transparency and human-centered design, understanding stakeholders' unique challenges, bridging communication gaps, and addressing ethical considerations for responsible AI deployment in healthcare settings [89, 49, 6].

6.6 Reliability and Accuracy

Benchmark	Size	Domain	Task Format	Metric
CD-CGPT[97]	100	Neuropathic Pain	Causal Inference	Precision, Recall
MediConfusion[79]	80,000	Radiology	Visual Question Answering (vqa)	Set accuracy, Individual accuracy
MEDALPACA[95]	160,000	Medicine	Question Answering	Accuracy
NLP-GC[56]	3,919	Genetics	Classification	Accuracy, F1 Score
SHE[98]	637	Social Sciences	Hypothesis-Evidence Relationship Classification	macro F1
PsyEval[99]	28,186	Mental Health	Question Answering	F1, AUC-ROC
LC-Stance[77]	400	Health	Stance Detection	F1-score
cMedKnowQA[100]	7,449	Medicine	Question Answering	H2

Table 4: This table presents a comprehensive overview of various benchmarks utilized in assessing AI models within healthcare and related domains. It includes details on the size, domain, task format, and evaluation metrics for each benchmark, providing insight into the diverse applications and performance measures employed in AI research.

The reliability and accuracy of AI predictions in healthcare are paramount, directly impacting clinical decision-making and patient outcomes. The integration of advanced predictive models, such as the XGB-DL model, has demonstrated a significant increase in the Area Under the Receiver Operating Characteristic (AUROC) curve by 13

The development of hybrid frameworks, such as those empowered by MLLMs, has been shown to improve reliability and accuracy while ensuring data security and promoting data sharing [90]. These frameworks are essential for maintaining the integrity of AI predictions and fostering trust in AI-driven healthcare solutions. Table 4 provides a detailed overview of representative benchmarks used to evaluate the reliability and accuracy of AI models in healthcare, highlighting the diversity of domains, task formats, and performance metrics.

Despite these advancements, significant limitations remain, particularly in models like ChatGPT, which have shown inconsistencies in answering causal questions, raising concerns about their reliability for causal inference [97]. Moreover, the evaluation of models in distinguishing between confusing image pairs has revealed performance levels below random guessing, underscoring the challenges in achieving reliable and accurate predictions in complex medical imaging tasks [79].

The LUNA model's effectiveness in detecting abnormal behaviors in LLMs emphasizes the critical need for reliability and accuracy in AI predictions within healthcare contexts [101]. Addressing these challenges is vital for ensuring AI technologies can be trusted to deliver consistent and accurate results, thereby enhancing patient care and clinical outcomes.

6.7 Regulatory and Compliance Issues

The deployment of AI in healthcare is subject to a complex regulatory and compliance landscape, posing significant challenges for integrating and operationalizing AI technologies. Regulatory frameworks, such as the EU General Data Protection Regulation (GDPR), impose stringent requirements on data privacy and security, necessitating robust measures to ensure compliance and protect patient information [88]. These regulations demand transparency in AI systems, critical for maintaining trust and accountability in healthcare applications.

A major regulatory challenge involves ensuring that AI models are both accurate and unbiased, as biases in training data can lead to skewed outcomes and potentially discriminatory practices in patient care [93]. Comprehensive evaluation frameworks assessing AI models' performance across diverse populations are essential for ensuring equitable healthcare solutions, addressing potential biases in language models that could affect AI-driven healthcare interventions' fairness and inclusivity [35].

Moreover, integrating AI technologies into existing healthcare systems must align with regulatory standards concerning medical device approval and safety. This involves rigorous testing and validation processes to ensure AI models meet necessary clinical efficacy and safety standards before deployment in clinical settings [3]. The complexity of these regulatory requirements can pose significant barriers to the swift adoption of AI solutions in healthcare, necessitating collaboration between developers, healthcare providers, and regulatory bodies to streamline compliance processes.

Handling sensitive patient data further complicates regulatory compliance, as AI systems must adhere to strict data protection laws to prevent unauthorized access and data breaches. Ensuring AI models do not inadvertently memorize or disclose private data is a critical concern that must be addressed through robust privacy protection mechanisms [91].

Navigating the regulatory and compliance challenges in deploying AI in healthcare requires a comprehensive approach that balances innovation with ethical and legal considerations. By effectively addressing the multifaceted challenges in patient-provider communication, particularly in high-stakes contexts like post-cancer treatment, the healthcare industry can leverage advanced AI technologies, such as LLMs, to enhance communication. This approach not only has the potential to improve patient outcomes but also ensures critical factors such as patient safety, privacy, and trust are upheld, fostering a more supportive and transparent healthcare environment [82, 89, 51, 49, 25].

7 Future Directions and Opportunities

7.1 Advancements in Multimodal and Multilingual Capabilities

Enhancing AI's multimodal and multilingual capabilities is crucial for advancing healthcare by facilitating comprehensive data processing and cross-cultural communication. Research should focus on developing sophisticated modality alignment techniques, improving model transparency, and establishing robust evaluation frameworks to ensure MLLMs' reliability in healthcare [6]. Expanding

AI-generated, for reference only.

datasets to encompass diverse medical scenarios and refining metrics are essential for mitigating hallucination issues and enhancing AI-driven healthcare accuracy [37].

Integrating LLMs into clinical settings necessitates validating benchmarks across varied clinical texts to effectively support decision-making and patient care [23]. Improving MLLMs' visual reasoning and addressing failure modes not covered by existing frameworks like MediConfusion is vital for robust AI healthcare applications [79].

In multilingual contexts, creating datasets in multiple languages and standardizing evaluation practices offer significant opportunities for healthcare progress. This involves exploring automated hyperparameter optimization and interdisciplinary LLM applications to enhance performance across AI domains [39]. Ethical considerations and guidelines for MLLMs in healthcare are crucial for responsible AI deployment [6].

Prioritizing these research areas will allow the healthcare industry to fully exploit AI's multimodal and multilingual capabilities, improving patient care and outcomes across diverse settings. Addressing downstream biases and exploring intersectional biases are also critical for ensuring equitable healthcare solutions [25].

7.2 Innovations in Training and Deployment Strategies

Innovating training and deployment strategies is essential for optimizing AI performance in healthcare. Future research should focus on hybrid methodologies that blend statistical and machine learning approaches, alongside evolving data complexity and analysis techniques [47]. Incorporating extensive patient data sequences into evaluation benchmarks like Asclepius can enhance their relevance and comprehensiveness [102].

Exploring additional metrics and deployment scenarios will increase benchmarks' adaptability across healthcare contexts [3]. Refining the TN-ML model to reduce human intervention and expanding its application to various diseases presents further innovation opportunities [8].

In therapeutic settings, controlled LLM integration is crucial for navigating complex emotional situations, thereby improving mental health interventions [36]. Future research should also refine the RAG approach to enhance document retrieval relevance, explore domain-specific applications, and improve AI-generated healthcare content accuracy [35].

Advancing predictive accuracy and reducing false positives through sophisticated models are vital for AI in diagnosing and managing chronic diseases [17]. Developing detection tools and policies to counter AI-generated misinformation is critical for maintaining AI-driven healthcare solutions' reliability [82].

By targeting these research directions, the healthcare industry can leverage AI technologies—such as LLMs and bandit algorithms—to improve patient-provider communication, streamline documentation, and personalize treatments. This approach addresses communication barriers and documentation burdens while utilizing patient-specific data for precision medicine, ultimately enhancing patient outcomes and healthcare efficiency [23, 49, 48].

8 Conclusion

The integration of Large Language Models (LLMs) and machine learning technologies in healthcare represents a significant advancement, particularly in the realm of risk prediction and management for chronic disease patients. The survey highlights the transformative potential of these technologies, emphasizing the importance of multimodal models and the need for continuous research to address current limitations. Challenges such as multilingual capabilities and the applicability of LLMs in low-resource languages remain areas for further development.

The findings underscore the critical role of LLMs in processing complex medical data, which facilitates personalized care and enhances clinical decision-making. Machine learning algorithms contribute to improved predictive accuracy, enabling the creation of tailored interventions that cater to individual patient needs. The survey also stresses the importance of addressing ethical considerations, such as bias and fairness, to ensure equitable access to AI-driven healthcare solutions.

Furthermore, the integration of AI technologies into healthcare systems enhances communication between providers and patients, supporting informed decision-making and improving patient outcomes. The nuanced role of AI in healthcare communication is evident, as perceptions of AI-generated messages vary depending on source disclosure. This survey underscores the need for ongoing innovation and collaboration among researchers, healthcare providers, and policymakers to fully leverage the potential of LLMs and machine learning in transforming healthcare delivery and improving patient outcomes.

AI-generated, for reference only.

References

[1] Nabeel Seedat, Caterina Tozzi, Andrea Hita Ardiaca, Mihaela van der Schaar, James Weatherall, and Adam Taylor. Unlocking historical clinical trial data with align: A compositional large language model system for medical coding, 2024.

[2] Pedro Henrique Paiola, Gabriel Lino Garcia, João Renato Ribeiro Manesco, Mateus Roder, Douglas Rodrigues, and João Paulo Papa. Adapting llms for the medical domain in portuguese: A study on fine-tuning and model evaluation, 2024.

[3] Yannis Bendi-Ouis, Dan Dutartre, and Xavier Hinaut. Deploying open-source large language models: A performance analysis, 2025.

[4] Yiheng Liu, Tianle Han, Siyuan Ma, Jiayue Zhang, Yuanyuan Yang, Jiaming Tian, Hao He, Antong Li, Mengshen He, Zhengliang Liu, Zihao Wu, Lin Zhao, Dajiang Zhu, Xiang Li, Ning Qiang, Dingang Shen, Tianming Liu, and Bao Ge. Summary of chatgpt-related research and perspective towards the future of large language models, 2023.

[5] Enshuo Hsu and Kirk Roberts. Leveraging large language models for knowledge-free weak supervision in clinical natural language processing, 2024.

[6] Qian Niu, Keyu Chen, Ming Li, Pohsun Feng, Ziqian Bi, Lawrence KQ Yan, Yichao Zhang, Caitlyn Heqi Yin, Cheng Fei, Junyu Liu, Benji Peng, Tianyang Wang, Yunze Wang, Silin Chen, and Ming Liu. From text to multimodality: Exploring the evolution and impact of large language models in medical practice, 2024.

[7] Armin Abdollahi, Negin Ashrafi, and Maryam Pishgar. Advanced predictive modeling for enhanced mortality prediction in icu stroke patients using clinical data, 2024.

[8] Yu-Jia An, Sheng-Chen Bai, Lin Cheng, Xiao-Guang Li, Cheng en Wang, Xiao-Dong Han, Gang Su, Shi-Ju Ran, and Cong Wang. Intelligent diagnostic scheme for lung cancer screening with raman spectra data by tensor network machine learning, 2023.

[9] Anas Maach, Jamila Elalami, Noureddine Elalami, and El Houssine El Mazoudi. An intelligent decision support ensemble voting model for coronary artery disease prediction in smart healthcare monitoring environments, 2022.

[10] Xin Tie, Muheon Shin, Ali Pirasteh, Nevein Ibrahim, Zachary Huemann, Sharon M. Castellino, Kara M. Kelly, John Garrett, Junjie Hu, Steve Y. Cho, and Tyler J. Bradshaw. Automatic personalized impression generation for pet reports using large language models, 2023.

[11] Ayomide Owoyemi, Emmanuel Nnaemeka, Temitope O. Benson, Ronald Ikpe, Blessing Nwachukwu, and Temitope Isedowo. Machine learning recommendation system for health insurance decision making in nigeria, 2023.

[12] Poorvesh Dongre, Majid Behravan, Kunal Gupta, Mark Billinghurst, and Denis Gračanin. Integrating physiological data with large language models for empathic human-ai interaction, 2024.

[13] Jessica Wilkins and Michael Rodriguez. Higher performance of mistral large on mmlu benchmark through two-stage knowledge distillation. 2024.

[14] Ernests Lavrinovics, Russa Biswas, Johannes Bjerva, and Katja Hose. Knowledge graphs, large language models, and hallucinations: An nlp perspective, 2024.

[15] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dan Mane, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viegas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. Tensorflow: Large-scale machine learning on heterogeneous distributed systems, 2016.

[16] Paul Bilokon, Oleksandr Bilokon, and Saeed Amen. A compendium of data sources for data science, machine learning, and artificial intelligence, 2023.

[17] Jamal Al-Karaki, Philip Ilono, Sanchit Baweja, Jalal Naghiyev, Raja Singh Yadav, and Muhammad Al-Zafar Khan. Predicting coronary heart disease using a suite of machine learning models, 2024.

[18] Sophia Krix, Ella Wilczynski, Neus Falgàs, Raquel Sánchez-Valle, Eti Yoles, Uri Nevo, Kuti Baruch, and Holger Fröhlich. Towards early diagnosis of alzheimer's disease: Advances in immune-related blood biomarkers and computational modeling approaches, 2023.

[19] Taewan Kim, Seolyeong Bae, Hyun Ah Kim, Su woo Lee, Hwajung Hong, Chanmo Yang, and Young-Ho Kim. Mindfuldiary: Harnessing large language model to support psychiatric patients' journaling, 2024.

[20] Donghoon Shin, Gary Hsieh, and Young-Ho Kim. Planfitting: Tailoring personalized exercise plans with large language models, 2023.

[21] Zhiyao Ren, Yibing Zhan, Baosheng Yu, Liang Ding, and Dacheng Tao. Healthcare copilot: Eliciting the power of general llms for medical consultation, 2024.

[22] Tao Feng, Lizhen Qu, Niket Tandon, Zhuang Li, Xiaoxi Kang, and Gholamreza Haffari. From pre-training corpora to large language models: What factors influence llm performance in causal discovery tasks?, 2024.

[23] Dave Van Veen, Cara Van Uden, Louis Blankemeier, Jean-Benoit Delbrouck, Asad Aali, Christian Bluethgen, Anuj Pareek, Malgorzata Polacin, Eduardo Pontes Reis, Anna Seehofnerova, Nidhi Rohatgi, Poonam Hosamani, William Collins, Neera Ahuja, Curtis P. Langlotz, Jason Hom, Sergios Gatidis, John Pauly, and Akshay S. Chaudhari. Adapted large language models can outperform medical experts in clinical text summarization, 2024.

[24] Yizhang Jin, Jian Li, Yexin Liu, Tianjun Gu, Kai Wu, Zhengkai Jiang, Muyang He, Bo Zhao, Xin Tan, Zhenye Gan, Yabiao Wang, Chengjie Wang, and Lizhuang Ma. Efficient multimodal large language models: A survey, 2024.

[25] Hanyao Huang, Ou Zheng, Dongdong Wang, Jiayi Yin, Zijin Wang, Shengxuan Ding, Heng Yin, Chuan Xu, Renjie Yang, Qian Zheng, and Bing Shi. Chatgpt for shaping the future of dentistry: The potential of multi-modal large language model, 2023.

[26] Katherine A. Rosenfeld, Maike Sonnewald, Sonia J. Jindal, Kevin A. McCarthy, and Joshua L. Proctor. Building understandable messaging for policy and evidence review (bumper) with ai, 2024.

[27] Jun-En Ding, Phan Nguyen Minh Thao, Wen-Chih Peng, Jian-Zhe Wang, Chun-Cheng Chug, Min-Chen Hsieh, Yun-Chien Tseng, Ling Chen, Dongsheng Luo, Chi-Te Wang, Pei fu Chen, Feng Liu, and Fang-Ming Hung. Large language multimodal models for 5-year chronic disease cohort prediction using ehr data, 2024.

[28] Junhao Zheng, Shengjie Qiu, Chengming Shi, and Qianli Ma. Towards lifelong learning of large language models: A survey, 2024.

[29] Phoebe Jing, Yijing Gao, Yuanhang Zhang, and Xianlong Zeng. Translating expert intuition into quantifiable features: Encode investigator domain knowledge via llm for enhanced predictive analytics, 2024.

[30] Zilong Wang, Nan Chen, Luna K. Qiu, Ling Yue, Geli Guo, Yang Ou, Shiqi Jiang, Yuqing Yang, and Lili Qiu. The potential and value of ai chatbot in personalized cognitive training, 2024.

[31] Roma Shusterman, Allison C. Waters, Shannon O'Neill, Phan Luu, and Don M. Tucker. An active inference strategy for prompting reliable responses from large language models in medical practice, 2024.

[32] Dmitry Scherbakov, Nina Hubig, Vinita Jansari, Alexander Bakumenko, and Leslie A. Lenert. The emergence of large language models (llm) as a tool in literature reviews: an llm automated systematic review, 2024.

[33] Marcos Fernández-Pichel, Juan C. Pichel, and David E. Losada. Search engines, llms or both? evaluating information seeking strategies for answering health questions, 2024.

[34] Hojjat Karami, David Atienza, and Anisoara Ionescu. Tee4ehr: Transformer event encoder for better representation learning in electronic health records, 2024.

[35] Aidan Gilson, Xuguang Ai, Thilaka Arunachalam, Ziyou Chen, Ki Xiong Cheong, Amisha Dave, Cameron Duic, Mercy Kibe, Annette Kaminaka, Minali Prasad, Fares Siddig, Maxwell Singer, Wendy Wong, Qiao Jin, Tiarnan D. L. Keenan, Xia Hu, Emily Y. Chew, Zhiyong Lu, Hua Xu, Ron A. Adelman, Yih-Chung Tham, and Qingyu Chen. Enhancing large language models with domain-specific retrieval augment generation: A case study on long-form consumer health question answering in ophthalmology, 2024.

[36] Zainab Iftikhar, Sean Ransom, Amy Xiao, and Jeff Huang. Therapy as an nlp task: Psychologists' comparison of llms and human peers in cbt, 2024.

[37] Jiawei Chen, Dingkang Yang, Tong Wu, Yue Jiang, Xiaolu Hou, Mingcheng Li, Shunli Wang, Dongling Xiao, Ke Li, and Lihua Zhang. Detecting and evaluating medical hallucinations in large vision language models, 2024.

[38] Nitin Rane, Saurabh Choudhary, and Jayesh Rane. Gemini versus chatgpt: applications, performance, architecture, capabilities, and implementation. *Journal of Applied Artificial Intelligence*, 5(1):69–93, 2024.

[39] Sen Huang, Kaixiang Yang, Sheng Qi, and Rui Wang. When large language model meets optimization, 2024.

[40] Adam White, Margarita Saranti, Artur d'Avila Garcez, Thomas M. H. Hope, Cathy J. Price, and Howard Bowman. Predicting recovery following stroke: deep learning, multimodal data and feature selection using explainable ai, 2023.

[41] Ethem Alpaydin. *Introduction to machine learning*. MIT press, 2020.

[42] Sophie Wharrie, Lisa Eick, Lotta Mäkinen, Andrea Ganna, Samuel Kaski, and FinnGen. Bayesian meta-learning for improving generalizability of health prediction models with similar causal mechanisms, 2024.

[43] Zihuai Zhao, Wenqi Fan, Jiatong Li, Yunqing Liu, Xiaowei Mei, Yiqi Wang, Zhen Wen, Fei Wang, Xiangyu Zhao, Jiliang Tang, and Qing Li. Recommender systems in the era of large language models (llms), 2024.

[44] Jackson A. Killian, Bryan Wilder, Amit Sharma, Daksha Shah, Vinod Choudhary, Bistra Dilkina, and Milind Tambe. Learning to prescribe interventions for tuberculosis patients using digital adherence data, 2019.

[45] Anuroop Sriram, Matthew Muckley, Koustuv Sinha, Farah Shamout, Joelle Pineau, Krzysztof J. Geras, Lea Azour, Yindalon Aphinyanaphongs, Nafissa Yakubova, and William Moore. Covid-19 prognosis via self-supervised representation learning and multi-image prediction, 2021.

[46] Rana Muhammad Shahroz Khan, Pingzhi Li, Sukwon Yun, Zhenyu Wang, Shahriar Nirjon, Chau-Wai Wong, and Tianlong Chen. Portllm: Personalizing evolving large language models with training-free and portable model patches, 2024.

[47] H Ij. Statistics versus machine learning. *Nat Methods*, 15(4):233, 2018.

[48] Yangyi Lu, Ziping Xu, and Ambuj Tewari. Bandit algorithms for precision medicine, 2021.

[49] Ziqi Yang, Xuhai Xu, Bingsheng Yao, Jiachen Li, Jennifer Bagdasarian, Guodong Gao, and Dakuo Wang. "i wish there were an ai": Challenges and ai potential in cancer patient-provider communication, 2024.

[50] Sue Lim and Ralf Schmälzle. The effect of source disclosure on evaluation of ai-generated messages: A two-part study, 2023.

[51] Yusif Ibrahimov, Tarique Anwar, and Tommy Yuan. Explainable ai for mental disorder detection via social media: A survey and outlook, 2024.

[52] Daniil Filienko, Yinzhou Wang, Caroline El Jazmi, Serena Xie, Trevor Cohen, Martine De Cock, and Weichao Yuwen. Toward large language models as a therapeutic tool: Comparing prompting techniques to improve gpt-delivered problem-solving therapy, 2024.

[53] Zelalem Gero, Chandan Singh, Yiqing Xie, Sheng Zhang, Praveen Subramanian, Paul Vozila, Tristan Naumann, Jianfeng Gao, and Hoifung Poon. Attribute structuring improves llm-based evaluation of clinical text summaries, 2024.

[54] Minbyul Jeong, Jiwoong Sohn, Mujeen Sung, and Jaewoo Kang. Improving medical reasoning through retrieval and self-reflection with retrieval-augmented large language models, 2024.

[55] Yi Luo, Zhenghao Lin, Yuhao Zhang, Jiashuo Sun, Chen Lin, Chengjin Xu, Xiangdong Su, Yelong Shen, Jian Guo, and Yeyun Gong. Ensuring safe and high-quality outputs: A guideline library approach for language models, 2024.

[56] Yujia Bao, Zhengyi Deng, Yan Wang, Heeyoon Kim, Victor Diego Armengol, Francisco Acevedo, Nofal Ouardaoui, Cathy Wang, Giovanni Parmigiani, Regina Barzilay, Danielle Braun, and Kevin S Hughes. Using machine learning and natural language processing to review and classify the medical literature on cancer susceptibility genes, 2019.

[57] Eric Lehman, Evan Hernandez, Diwakar Mahajan, Jonas Wulff, Micah J. Smith, Zachary Ziegler, Daniel Nadler, Peter Szolovits, Alistair Johnson, and Emily Alsentzer. Do we still need clinical language models?, 2023.

[58] Bhawesh Kumar, Jonathan Amar, Eric Yang, Nan Li, and Yugang Jia. Selective fine-tuning on llm-labeled data may reduce reliance on human annotation: A case study using schedule-of-event table detection, 2024.

[59] Jiageng Wu, Xian Wu, Yefeng Zheng, and Jie Yang. Medkp: Medical dialogue with knowledge enhancement and clinical pathway encoding, 2024.

[60] Zhijie Bao, Wei Chen, Shengze Xiao, Kuang Ren, Jiaao Wu, Cheng Zhong, Jiajie Peng, Xuanjing Huang, and Zhongyu Wei. Disc-medllm: Bridging general large language models and real-world medical consultation, 2023.

[61] Chanseo Lee, Sonu Kumar, Kimon A. Vogt, Sam Meraj, and Antonia Vogt. Advancing complex medical communication in arabic with sporo arasum: Surpassing existing large language models, 2024.

[62] Jeremie Bogaert and Francois-Xavier Standaert. A question on the explainability of large language models and the word-level univariate first-order plausibility assumption, 2024.

[63] Daniel P. Jeong, Pranav Mani, Saurabh Garg, Zachary C. Lipton, and Michael Oberst. The limited impact of medical adaptation of large language and vision-language models, 2025.

[64] Hanguang Xiao, Feizhong Zhou, Xingyue Liu, Tianqi Liu, Zhipeng Li, Xin Liu, and Xiaoxuan Huang. A comprehensive survey of large language models and multimodal large language models in medicine, 2024.

[65] Bell Raj Eapen, Kamran Sartipi, and Norm Archer. Serverless on fhir: Deploying machine learning models for healthcare on the cloud, 2020.

[66] Scott L. Fleming, Alejandro Lozano, William J. Haberkorn, Jenelle A. Jindal, Eduardo P. Reis, Rahul Thapa, Louis Blankemeier, Julian Z. Genkins, Ethan Steinberg, Ashwin Nayak, Birju S. Patel, Chia-Chun Chiang, Alison Callahan, Zepeng Huo, Sergios Gatidis, Scott J. Adams, Oluseyi Fayanju, Shreya J. Shah, Thomas Savage, Ethan Goh, Akshay S. Chaudhari, Nima Aghaeepour, Christopher Sharp, Michael A. Pfeffer, Percy Liang, Jonathan H. Chen, Keith E. Morse, Emma P. Brunskill, Jason A. Fries, and Nigam H. Shah. Medalign: A clinician-generated dataset for instruction following with electronic medical records, 2023.

[67] Yusheng Liao, Yutong Meng, Hongcheng Liu, Yanfeng Wang, and Yu Wang. An automatic evaluation framework for multi-turn medical consultations capabilities of large language models, 2023.

[68] Andreas Holzinger, Markus Plass, Katharina Holzinger, Gloria Cerasela Crisan, Camelia-M. Pintea, and Vasile Palade. A glass-box interactive machine learning approach for solving np-hard problems with the human-in-the-loop, 2017.

[69] Alexander Jung and Pedro H. J. Nardelli. An information-theoretic approach to personalized explainable machine learning, 2020.

[70] Liam Hazan, Gili Focht, Naama Gavrielov, Roi Reichart, Talar Hagopian, Mary-Louise C. Greer, Ruth Cytter Kuint, Dan Turner, and Moti Freiman. Leveraging prompt-learning for structured information extraction from crohn's disease radiology reports in a low-resource language, 2024.

[71] Francisco Y Cai, Chengsheng Mao, and Yuan Luo. Constrained tensor factorization for computational phenotyping and mortality prediction in patients with cancer, 2021.

[72] L. Julian Lechuga Lopez, Tim G. J. Rudner, and Farah E. Shamout. Informative priors improve the reliability of multimodal clinical data classification, 2023.

[73] Hongwei Jin, George Papadimitriou, Krishnan Raghavan, Pawel Zuk, Prasanna Balaprakash, Cong Wang, Anirban Mandal, and Ewa Deelman. Large language models for anomaly detection in computational workflows: from supervised fine-tuning to in-context learning, 2024.

[74] Elias Chaibub Neto, Phil Snyder, Solveig K Sieberts, and Larsson Omberg. Stable predictions for health related anticausal prediction tasks affected by selection biases: the need to deconfound the test set features, 2020.

[75] Thokozile Manaka, Terence Van Zyl, Alisha N Wade, and Deepak Kar. Using machine learning to fuse verbal autopsy narratives and binary features in the analysis of deaths from hyperglycaemia, 2022.

[76] Tammy Jiang, Jaimie L Gradus, and Anthony J Rosellini. Supervised machine learning: a brief primer. *Behavior therapy*, 51(5):675–687, 2020.

[77] Joseph Gatto, Madhusudan Basak, Yash Srivastava, Philip Bohlman, and Sarah M. Preum. Scope of large language models for mining emerging opinions in online health discourse, 2024.

[78] Wei Zhu, Wenfeng Li, Xing Tian, Pengfei Wang, Xiaoling Wang, Jin Chen, Yuanbin Wu, Yuan Ni, and Guotong Xie. Text2mdt: Extracting medical decision trees from medical texts, 2024.

[79] Mohammad Shahab Sepehri, Zalan Fabian, Maryam Soltanolkotabi, and Mahdi Soltanolkotabi. Mediconfusion: Can you trust your ai radiologist? probing the reliability of multimodal medical foundation models, 2024.

[80] Yutong Hu, Quzhe Huang, Mingxu Tao, Chen Zhang, and Yansong Feng. Can perplexity reflect large language model's ability in long text understanding?, 2024.

[81] Gloria Ashiya Katuka, Alexander Gain, and Yen-Yun Yu. Investigating automatic scoring and feedback using large language models, 2024.

[82] Luigi De Angelis, Francesco Baglivo, Guglielmo Arzilli, Gaetano Pierpaolo Privitera, Paolo Ferragina, Alberto Eugenio Tozzi, and Caterina Rizzo. Chatgpt and the rise of large language models: the new ai-driven infodemic threat in public health. *Frontiers in public health*, 11:1166120, 2023.

[83] Ruchik Mishra, Karla Conn Welch, and Dan O Popa. Human-mediated large language models for robotic intervention in children with autism spectrum disorders, 2024.

[84] Wenli Zhang, Jiaheng Xie, Zhu Zhang, and Xiang Liu. Depression detection using digital traces on social media: A knowledge-aware deep learning approach, 2023.

[85] Jun Liu, Chaoyun Zhang, Jiaxu Qian, Minghua Ma, Si Qin, Chetan Bansal, Qingwei Lin, Saravan Rajmohan, and Dongmei Zhang. Large language models can deliver accurate and interpretable time series anomaly detection, 2024.

[86] Rongsheng Wang, Yaofei Duan, ChanTong Lam, Jiexi Chen, Jiangsheng Xu, Haoming Chen, Xiaohong Liu, Patrick Cheong-Iao Pang, and Tao Tan. Ivygpt: Interactive chinese pathway language model in medical domain, 2023.

[87] Jeffery L. Painter, Venkateswara Rao Chalamalasetti, Raymond Kassekert, and Andrew Bate. Automating pharmacovigilance evidence generation: Using large language models to produce context-aware sql, 2024.

[88] Nadia Burkart and Marco F Huber. A survey on the explainability of supervised machine learning. *Journal of Artificial Intelligence Research*, 70:245–317, 2021.

[89] Q. Vera Liao and Jennifer Wortman Vaughan. Ai transparency in the age of llms: A human-centered research roadmap, 2023.

[90] Cheng Su, Jinbo Wen, Jiawen Kang, Yonghua Wang, Yuanjia Su, Hudan Pan, Zishao Zhong, and M. Shamim Hossain. Hybrid rag-empowered multi-modal llm for secure data management in internet of medical things: A diffusion-based contract approach, 2024.

[91] Yuqi Yang, Xiaowen Huang, and Jitao Sang. Exploring the privacy protection capabilities of chinese large language models, 2024.

[92] Heinrich Peters and Sandra Matz. Large language models can infer psychological dispositions of social media users, 2024.

[93] Abel Salinas, Louis Penafiel, Robert McCormack, and Fred Morstatter. "im not racist but...": Discovering bias in the internal knowledge of large language models, 2023.

[94] Ali Borji. A categorical archive of chatgpt failures, 2023.

[95] Tianyu Han, Lisa C Adams, Jens-Michalis Papaioannou, Paul Grundmann, Tom Oberhauser, Alexander Löser, Daniel Truhn, and Keno K Bressem. Medalpaca–an open-source collection of medical conversational ai models and training data. *arXiv preprint arXiv:2304.08247*, 2023.

[96] Shibin Wu, Bang Yang, Zhiyu Ye, Haoqian Wang, Hairong Zheng, and Tong Zhang. Improving medical report generation with adapter tuning and knowledge enhancement in vision-language foundation models, 2023.

[97] Ruibo Tu, Chao Ma, and Cheng Zhang. Causal-discovery performance of chatgpt in the context of neuropathic pain diagnosis, 2023.

[98] Sai Koneru, Jian Wu, and Sarah Rajtmajer. Can large language models discern evidence for scientific hypotheses? case studies in the social sciences, 2024.

[99] Mihael Arcan, David-Paul Niland, and Fionn Delahunty. An assessment on comprehending mental health through large language models, 2024.

[100] Haochun Wang, Sendong Zhao, Zewen Qiang, Zijian Li, Nuwa Xi, Yanrui Du, MuZhen Cai, Haoqiang Guo, Yuhan Chen, Haoming Xu, Bing Qin, and Ting Liu. Knowledge-tuning large language models with structured medical knowledge bases for reliable response generation in chinese, 2023.

[101] Da Song, Xuan Xie, Jiayang Song, Derui Zhu, Yuheng Huang, Felix Juefei-Xu, and Lei Ma. Luna: A model-based universal analysis framework for large language models, 2024.

[102] Wenxuan Wang, Yihang Su, Jingyuan Huan, Jie Liu, Wenting Chen, Yudi Zhang, Cheng-Yi Li, Kao-Jung Chang, Xiaohan Xin, Linlin Shen, and Michael R. Lyu. Asclepius: A spectrum evaluation benchmark for medical multi-modal large language models, 2024.

Disclaimer:

SurveyX is an AI-powered system designed to automate the generation of surveys. While it aims to produce high-quality, coherent, and comprehensive surveys with accurate citations, the final output is derived from the AI's synthesis of pre-processed materials, which may contain limitations or inaccuracies. As such, the generated content should not be used for academic publication or formal submissions and must be independently reviewed and verified. The developers of SurveyX do not assume responsibility for any errors or consequences arising from the use of the generated surveys.