

Analysis of Artificial Intelligence in the Healthcare Sector

Yash Yadav

Department of Commerce

Maharaja Agrasen college university of delhi

1. INTRODUCTION

The intersection of artificial intelligence and health care is one of the most radical technological revolutions of the twenty-first century, completely changing the paradigm of diagnostics, treatment procedure, and the overall structure of the healthcare delivery system across the world. This crossroads did not emerge as a spontaneous arrival but through decades of gradual technological progress, the changes in computational abilities, and an ever-digitized healthcare system that produces enormous amounts of data, which can be analyzed by algorithms. The modern environment of AI-powered healthcare is built on the groundwork of the contributions made in the 1970s and 1980s when the first ideas of expert systems that can simulate clinical decision-making were developed by computer scientists and medical professionals.

Artificial intelligence in medical use dates back to the primitive rule-based expert systems, like the 1970s MYCIN, developed at Stanford University, which showed that computer programs could perform at the level of an expert, in diagnosing blood infections and recommending antibiotic doses (Shortliffe, 2012). It was a seminal work that laid the conceptual basis on the fact that medical knowledge could be encoded in a computation system and be used to solve clinical problems with logical reasoning. Yet, these early systems worked in very limited scopes and were based on manually designed rules which were not resistant facing the complexity and variability of the medical practice in the real world. These constraints of the rule-based methods all became more and more evident during the 1990s and 1980s, as the medical community realized that healthcare decision-making is a matter of probabilistic thought, pattern recognition and situational consideration beyond the scope of the simple if-then logical constructions.

The technological trends initiating AI advancement in healthcare accelerated dramatically at the start of the twenty-first century, driven by three convergent technological forces: the exponential growth in the computational processing power as defined by the Moore's Law, the spread of digital health records producing unprecedented amounts of structured and unstructured medical data, and the underlying advances in machine learning algorithms: specifically, the advent of deep learning architectures inspired by neural network research dating back to the 1940s but now capable of running computationally (LeCun et al., 2015). It was a paradigmatic shift, when the use of machine learning based on data began to replace rule-based expert systems, the AI systems started to find patterns and relationships in medical data without explicit programming, learning from examples rather than following predetermined rules.

The scope of modern healthcare AI applications is truly impressive and covers numerous clinical and non-clinical areas. Convolutional neural networks have also been shown to be able to interpret medical imaging with human-level or better accuracy in radiology, pathology, dermatology, and ophthalmology, with landmark studies demonstrating AI systems at radiologist-level accuracy in diagnosing pneumonia through chest X-rays, predicting diabetic retinopathy through fundus photographs, and classifying skin lesions with dermatologist-level accuracy (Esteva et al., 2017; Rajpurkar et al., 2017). The proposed predictive analytics based on the use of data on electronic health records can help in identifying patients at risk of sepsis, hospital readmission, and disease advancement earlier, which might help alter the current reactive healthcare paradigm toward the preventive paradigm. Natural language processing systems can identify clinical information in unstructured physician notes, radiology reports, and medical literature, and chatbots can assist the initial evaluation of symptoms and triage a patient. Surgical robots with AI capabilities increase the accuracy of surgical procedures, minimize the complications of the operation and allow conducting a minimum of invasive surgery. Machine learning is increasingly becoming part of drug discovery pipelines to discover therapeutic targets, predict molecular

properties, and optimize the design of clinical trials, which could shorten discovery timelines and cut the cost of pharmaceutical R&D (Chen et al., 2018).

Although these technological milestones exist and there is a significant amount of hype about the importance of artificial intelligence in helping to solve long-standing issues in healthcare, such as clinical burnout, misdiagnoses, healthcare access inequalities, and rising healthcare costs, the adoption of artificial intelligence in healthcare delivery faces complex barriers on both technical, ethical, regulatory, financial, and sociocultural levels. The issue domain that this dissertation will discuss is the multi-layered nature of the opportunities and challenges that the introduction of AI technology presents the modern healthcare system through an interdisciplinary lens that includes technological, clinical, ethical, and commercial viewpoints that would be relevant to the commerce students becoming acquainted with the new field of study.

Technically, AI systems have inherent issues associated with data quality, algorithm transparency, and overall generalizability. Healthcare data are inherently complex and heterogeneous, and frequently low-quality standards, and electronic health records differ significantly among institutions in term and structure and completeness (Rajkumar et al., 2018). The models developed by machine learning on the data of particular groups of people or healthcare facilities might not work in other contexts, which is known as dataset drift or external invalidity, which has led to ethical and legal issues about algorithmic bias and health equity. Interpretability issues in complex deep learning systems are especially severe with the black box nature of advanced neural network models, which are especially important in medical settings where medical judgements require coherent explanations and responsibility. As an AI system suggests a particular diagnosis or treatment, both clinicians and patients have a legitimate right to understand the reasons in terms of understandable medical reasoning as opposed to being obscured through statistical relationships in a high-dimensional feature space.

The AI healthcare applications are touched with ethical considerations, which include questions of patient privacy, informed consent, and algorithmic fairness, and the right balance between human judgment and machine suggestions. Training AI systems also needs access to sensitive medical information, with some conflicting tensions between the societal gains of better algorithms and personal privacy rights under laws like the Health Insurance Portability and Accountability Act (HIPAA) in the United States or the General Data Protection Regulation (GDPR) in the European Union. Recorded cases of algorithmic bias, in which AI systems perform differently when exposed to demographic groups based on race, ethnicity, gender, or socioeconomic status, bring significant questions of equity, since healthcare AI has an opportunity to reproduce or exacerbate existing health inequalities instead of reducing them (Obermeyer et al., 2019). Ethical models of human clinical research, which have been developed and refined over decades due to the historical abuses of this field, need to be adjusted to new questions arising in the work of AI systems that keep learning and improving after implementation.

Regulatory and governance-wise, healthcare systems across the globe are struggling with the need to develop proper regulatory frameworks places on AI technologies that do not necessarily fall under the conventional medical device classification patterns. The regulative problem of AI algorithms that are trained on continuous learning is unlike that of traditional medical devices with fixed specifications- does each update of an algorithmic specification need a new regulatory clearance, and how can safety be guaranteed in a dynamical change over time? The regulatory agencies such as the United States Food and Drug Administration (FDA), the European Medicines Agency (EMA), and the national health authorities in the countries, including the United Kingdom, Canada, and Australia, have started to work out AI-specific regulatory pathways, but there are still significant doubts about the standards of validation, performance monitoring, and post-market surveillance (Gerke et al., 2020). The business aspects of AI in medicine introduce no less complicated concerns especially when it comes to students and practitioners of the business, business, and healthcare management sectors. The market of artificial intelligence in healthcare has been growing at an explosive pace all over the world, with market estimates predicting the valuations of tens of billions of dollars by the mid-2020s and attracting billions of venture capital investment and spawning the creation of dozens of health technology startups alongside efforts by established technology giants like Google, Microsoft, Amazon, and IBM (Topol, 2019). This commercialization begs the question about what the proper balance should be between the incentives to innovate and the equitable access especially with healthcare being a social good where market failures should be rectified by the public sector. The monetization of AI technologies requires significant capital expenditures on data infrastructure, computational power,

and human workforce education, which may further widen the gap between the healthcare organizations, which are well-capacitated in the high-income nations, and those, which are resource constrained and have the most urgent healthcare demands.

The organizational and workforce impact of AI application in healthcare is not limited to the technological aspects as it involves the professional identity, clinical practices, and future medical workforce. Medical practitioners, nurses, and other health care providers are afraid of deskilling, professional autonomy, and possible job displacement, but they also feel burnout due to the burden of administration that AI may help eliminate. Existing concepts of successful AI integration demand not the deployment of technologies but sociotechnical change, involving the consideration of human factor and management of the organization change and the development of augmented intelligence paradigms in which the use of AI will complement rather than substitute human clinical judgment (Shortliffe and Sepulveda, 2018). The COVID-19 pandemic occurred in late 2019 and is still ongoing into early 2020s and became both an impetus and a trial of AI healthcare applications, boosting the use of telemedicine, predictive analytics of resources distribution, and AI-enabled drug repurposing projects. The pandemic also highlighted shortcomings and weaknesses as AI systems trained on pre-pandemic data were ill-equipped to handle redistribution and new clinical manifestations, which highlights the need to ensure robust and flexible systems that can be deployed in unprecedented conditions (Wynants et al., 2020).

It is placed in the context of this dynamic, multifaceted landscape, where this dissertation provides an in-depth examination of artificial intelligence within the healthcare field due to the analytical approach to the scholarly literature on the subject in the technological, clinical, ethical, regulatory, and commercial spheres. The study solves the basic questions: What are the present potential of AI technologies and their drawbacks in different healthcare applications? What is the perception and experience of AI integration by the various stakeholder groups, such as patients, clinicians, policymakers and commercial entities? What ethical systems and regulatory strategies should provide the optimal balance between innovation and patient safety and equity? How do we use business models and economic structures to enable sustainable and equitable AI implementation? What are the major gaps in the existing knowledge, and in which directions do future studies need to be oriented?

The literature-based study is an addition to the knowledge as it has summarized various bodies of research, which are usually in silos due to their limited scope in different disciplines, giving a holistic view of it, especially to students of commerce who will define business, policy, and management aspects of the further development of AI in healthcare. Through a critical review of the available evidence, research gaps and suggestions on frameworks that can be used to explain the multifaceted implications of this transformative technology, the work will inform more effective, equitable and ethically sound approaches to using artificial intelligence to achieve better health outcomes and improved healthcare delivery systems across the world. The analysis acknowledges that to achieve the potential of AI, it is not only necessary to have technological sophistication but also an attentiveness to the social, organizational, economical, and ethical environments in which AI technologies work-environments that ultimately determine whether AI becomes a source of reducing or increasing health disparities, empowering or displacing medical practitioners, and democratizing or concentrating access to advanced medical functions. The general aim of this dissertation is to offer a critical, analytical and rigorous examination of artificial intelligence usage in the healthcare sector by conducting a systematic review and synthesis of the available academic literature. In particular, the work will chart the terrain of the existing body of knowledge on AI healthcare technologies, find patterns and trends in research and practice, critically analyze statements about the opportunities and constraints of AI, reflect on the ethical and legal issues that come with the implementation of AI, discuss the business and economic aspects of healthcare AI-based industries, and finally provide some practical recommendations on how researchers, policymakers, healthcare managers, and business leaders to navigate this fast-growing sphere.

2. STATEMENT OF PROBLEM

The incorporation of artificial intelligence (AI) into healthcare systems has become a new paradigm, but it has struggled to cope with complex issues within the technical, ethical, regulatory, economic, and implementation fields. These obstacles, which continue to increase its investments and scientific attention within the last ten years, require hard interdisciplinary evaluation in order to outline the proper regulatory role of AI and its social consequences.

An essential problem, which is rather technical, is the generality and strength of AI models. On the one hand, these systems perform well on curated datasets, but on the other hand, they fail in real-world clinical setting due to change of distribution in patient demographics, equipment, or disease manifestation (Zech et al., 2018; Nagendran et al., 2020). This is due to poor sampling of training data, biased procedures of evaluation and using spurious correlations instead of clinically significant features weaken their use in patient care.

Intimately connected is an algorithmic bias that worsens health inequity. AI also has disproportionate and unequal performance based on racial, ethnic, gender, and socioeconomic backgrounds, continuing the past discriminations inherent in datasets (Obermeyer et al., 2019; Gianfrancesco et al., 2018). Mechanisms are lopsided training statistics and optimization to aggregate accuracy in place of equity. It needs the mitigation of lifecycle bias, fairness metrics, and institutional commitments to avoid the continued expansion of inequalities among the vulnerable populations. Another challenge is the black-box models, particularly the deep neural networks, being opaque. Compared to the transparent clinical reasoning, these systems provide predictions that cannot be explained with rational grounding, which makes it difficult to trust and hold accountable and without integration (Rudin, 2019; Tonekaboni et al., 2019). Post-hoc explanations can be very untrustworthy or unreadable and require a trade-off between precision and understandability without any definite rules that guide the usage in clinical situations.

Privacy and consent dilemmas are ethical issues due to the use of AI which requires massive sensitive data. The classical models fail in any machine learning scenarios where data is used in aggregate, multipurpose modes (Price and Cohen, 2019). Societal advantages and personal privacy remain in conflict, and methods such as federated learning may be used to partially resolve the conflict between the two with the accuracy-privacy trade-off. Commercialization is also complicated by the ownership of data and fair share of benefits.

The current regulatory frameworks are far behind the progress of AI, and they do not fit well into an adaptive system that learns after being deployed (Gerke et al., 2020; FDA, 2019). Among those that are yet to be resolved are evidence standards, bias evaluations, transparency mandates, and global harmonization, which has led to a disjointed ecosystem that could potentially cause stifled innovations as well as unsafe applications.

Economically, the issues of sustainability are unpredictable reimbursements, prohibitive prices, and incentive mismatches, which may cause underinvestment in underserved disciplines and market monopoly (Hwang et al., 2019). This is further exacerbated by implementation gaps, which disrupt sociotechnical barriers such as workflow disruptions, usability problems, and resistance to obstruct the realisation of clinical value (Sendak et al., 2020).

The implications of workforce include deskilling, loss of autonomy, and liability, and require the development of advanced medical education (Blease et al., 2019). Literature is also biased, has defective validations, and distorted publications, which reduce evidence synthesis (Roberts et al., 2021).

Taken together, these interactive issues require adaptive governance. This dissertation is a literature review that analyzes the assertions, gaps, and offers fair frameworks, which commerce professionals need to address when negotiating the commercialization of AI in healthcare.

3. SCOPE OF RESEARCH

The present scope of this dissertation is delineated but comprehensive in nature and focused on the use of artificial intelligence (AI) in healthcare domains and is undertaken through a systematic review of academic literature in the majority of cases covering the period 2000 to 2025. The assessment of specific inclusion and exclusion criteria allows the study to be analytically manageable without sacrificing the coherence in the multidimensional aspects of technological changes, integration in the clinical sphere, and implications on society.

The timeframe mostly refers to the time since 2000 to 2025, the span of which embraces the continuum of machine learning and deep learning ascent in healthcare settings. This period is marked by the transition of primitive rule-based expert systems to powerful data-driven paradigms, such as deep learning boom in the 2010s, and further developments up to 2025, such as regulatory framework transformation, the development of ethical discourses, and the extensive effects of the COVID-19 pandemic. Classical investigations between the 1970s and 1990s are selectively integrated to place historical trends within historical context, but recent literature is favoured as the systematic review to ensure relevance to the existing transformative processes.

The inquiry has a spectrum of technologies related to AI and machine learning modalities to healthcare, such as supervised learning to make diagnostic predictions and risk assessments, deep learning architectures, such as convolutional neural networks, to analyse images and transformers to analyse natural language, unsupervised to detect patterns, reinforcement learning to optimize therapy, computer vision to guide surgery, and predictive analytics based on electronic health records. Purely statistical methods that lack learning components, more conventional rule-based decision support systems, and more general health information technologies are excluded, narrowing the focus to intelligent, adaptive systems.

Clinically and application-wise, the dissertation covers AI implementation in diverse health care settings, such as diagnostic specialties, such as radiology and pathology to classify diseases, predictive models to forecast sepsis and allocate resources, therapeutic development in personalized medicine and drug discovery, surgical advancements through robotics and intraoperative navigation, mental health interventions through chatbots and risk prediction, operational efficiencies in scheduling and patient management, and the use of AI in epidemiological modeling of the population. They are used across a wide variety of care environments, such as acute hospital and primary clinic and consumer-facing environments, and the views are taken based on the high-income and resource-limited global settings in both high-income and resource-constrained environments to highlight equity issues.

They incorporate the perspectives of stakeholders to enhance the interdisciplinary understanding, including the experience of clinical professionals with workflow integration and trust, patient thought of access and inequalities, regulatory focus on governance and safety, commercial evaluation of market feasibility and investment, ethical reflection of bias and privacy, and policy reflection of reimbursement and labor changes. This interdisciplinary lens is based on a variety of fields, such as computer science, clinical medicine, bioethics, health economics, regulatory law, and implementation science, and enabled the work to be more holistic when compared to siloed approaches to analysis, and provide a comprehensive value to commerce students who interact with the business and managerial aspects of AI.

The study is based on the literature synthesis methodology, not the primary empirics, using the articles in peer-reviewed journals and systematic reviews, policy reports, and some gray literature by the authorities, avoiding original data collection or trials. Geographically, although inclusive of international contributions, the literature composes of most developments in wealthy countries such as the United States, Europe, East Asia, and this has raised serious questions of how this may affect global health equity, and new uses by low-income communities.

The dissertation acknowledges intrinsic constraints, not conducting primary research investigations, exploring in depth the algorithmic mathematics, long-term outcome studies, exposing proprietary system code or writing in languages other than English, which would possibly have missed non-Western wisdom. These constraints also lead to providing a critically synthesized evaluation of the AI potential, its challenges, and its way forward in the healthcare sector to inform scholars, policymakers, and professionals working in this critical field.

4. RESEARCH QUESTIONS

This dissertation was guided by seven research questions:

RQ1: What are the existing technological possibilities and uses of AI systems in healthcare?

RQ2: What do we know about the actual clinical efficacy, patient outcomes, and effects of AI on the health system in the real sense?

RQ3: What are the ethical issues, forms of bias, and health equity concerns of AI implementation in the medical field?

RQ4: What regulatory frameworks are used by major jurisdictions to deal with AI in healthcare?

RQ5: What are the business models and market dynamics of the AI in healthcare industry, commercial dimensions?

RQ6: What are the barriers to the successful implementation of AI in healthcare systems, and what are the implementation challenges?

RQ7: What are the major gaps in the research concerning AI in healthcare?

5. OBJECTIVES

The key goal of this dissertation was to perform thorough literature-based research of artificial intelligence in healthcare field, including technological, clinical, ethical, regulatory, and commercial aspects. To achieve this goal the following specific objectives directed the research:

1. To conduct a systematic literature review and synthesize of the available scholarly literature regarding the use of artificial intelligence in various fields of healthcare.
2. To explore thoroughly the ethical, legal, and regulatory issues related to the deployment of artificial intelligence in the healthcare environment.
3. To examine the commercial, economic, and market aspects of artificial intelligence in healthcare using non-disciplinary cross-interdisciplinary approaches that combine commerce, business strategy, and health economics.
4. To reveal and critically review substantial gaps in the current research on AI in healthcare.
5. To develop conclusions and recommendations to be applied to various groups of stakeholders.

6. METHODOLOGY:

This dissertation used the systematic literature review methodology, based on PRISMA guidelines, that relies on narrative synthesis techniques and critical appraisal strategies. The methodology was based on a post-positivist epistemological position and understanding that objective knowledge could be built in the course of systematic research but also admitting the fact that ultimate objectivity was impossible and that several acceptable interpretations could co-exist.

Information Sources and Search Strategy: Multiple databases were systematically searched that covered in depth interdisciplinary literature: PubMed/MEDLINE (biomedical literature), Scopus and Web of Science (multidisciplinary scientific literature), IEEE Xplore (computer science/engineering), Business Source Complete (business/management), and Google Scholar (comprehensive multidisciplinary coverage). Additional sources were Cochrane Library, preprint archives (arXiv, medRxiv, bioRxiv), regulatory databases (FDA, EMA) as well as gray literature in WHO, national health agencies, and health technology assessment agencies. The search terms were a combination of AI/ML terms (artificial intelligence, machine learning, deep learning, neural networks) and healthcare terms (diagnosis, treatment, clinical, medical imaging) and the areas of interest (bias, fairness, effectiveness, implementation), depending on the syntax of each database. Searches were done on 2000-2025 with selective inclusion of earlier foundational works, which included English-language publications.

Eligibility criteria: The articles met the following criteria: The article should cover AI/ML applications in healthcare, be a peer-reviewed publication or a credible gray literature, use any form of empirical study or theoretical analysis, and be within 2000-2025, in English, and in full text. The exclusions were materials with some irrelevant topic, abstracts, but no full text, low-quality publications, and low-quality publications, and predatory publications.

Selection Process: The process followed was five-stage and included: (1) systematic database search which included removal of duplicates, (2) filtering of title/abstracts by eligibility criteria, (3) full-text screening of passing citations, (4) backward/forward citation searching and (5) inclusion finalization.

Synthesis and Analysis: Since there was heterogeneity in various study design and domains, narrative thematic synthesis was used as an alternative to quantitative meta-analysis. Evidence was categorized into thematic areas based on research questions (technical capabilities, clinical effectiveness, ethics/bias, regulation, commercial dimensions, implementation, research gaps) as well. In both domains, the range of evidence was described in terms of narrative synthesis, with identified patterns and contradictions, along with a summary of key findings, and the strength of the evidence. Validity, tensions, assumptions, whose perspectives are represented, contextualized findings and generalizability were critical analyzed. The synthesis directly involved interdisciplinary knowledge (medicine, computer science, ethics, economics, policy) and systematically reported gaps in the research.

7. LITERATURE REVIEW/ FINDINGS

The body of artificial intelligence research in healthcare has grown exponentially in the last 20 years as a result of the accelerated pace of technological change in machine learning techniques and the augmented understanding of the transformative potential of artificial intelligence in clinical, operational, and research settings. The overall review under discussion is a syntactic compilation of academic debate divided into themes in order to shed light on critical investigation, discussion, and implementation. They include AI applications in diagnostic and clinical decision-making, therapeutic interventions and personalized medicine, operational and administrative optimization, ethical and societal issues, regulatory and governance frameworks, economic and commercial aspects and implementation challenges and factors leading to success. This topical structure helps to comprehend not only the progress within the sphere but also those transverse issues that define the healthcare AI environment.

7.1 Historical Evolution and Foundational Concepts

The conceptual roots of artificial intelligence in medicine can be traced to the infantile years of computer science in the mid-twentieth century when scientists first thought of how computational systems might mimic expert human thought in the specialized fields. The original 1976 publication of MYCIN, an expert system to diagnose bacterial infections and prescribe antimicrobial therapy, was the first such publication to demonstrate that rule-based systems with medical knowledge encoded would be able to perform at the level of a human expert. MYCIN used about 600 production rules based on infectious disease professionals, which used backward chaining reasoning over diagnostic hypotheses and treatment suggestions. In spite of its spectacular success in the controlled tests, diagnosing infections with nearly 69 percent accuracy, as compared to 80 percent by human experts in current research, the clinical implementation of MYCIN was never realized, due to its computational limitations, difficulties in integrating the system with clinical processes and physician resistance to follow the computer-generated suggestions (Buchanan and Shortliffe, 1984).

The following decades saw the spread of medical expert systems such as INTERNIST-I in internal medicine diagnosis, CADUCEUS in complex diagnostic reasoning, and DXplain as a diagnostic decision support system but the clinical impact was not as great. Miller (1994) gave critical analysis of limitations of expert systems, which he stated were brittle when faced with unusual cases, lacked experience learning, maintenance because of the changing medical knowledge, and the “knowledge acquisition bottleneck” whereby medical knowledge extraction and encoding were extremely labor-intensive. These constraints helped drive what Szolovits (2019) described as an AI winter in medical computing of the 1990’s as interest in the field faded due to the acknowledgment that rule-based systems would not be scalable to meet the complexity and subtlety of clinical practice in the real world.

The revival of AI in healthcare occurred at the beginning of the twenty-first century and was driven by the confluence of several enabling factors. LeCun et al. (2015) explained that breakthroughs in deep learning, especially convolutional neural networks in image recognition and recurrent neural networks in sequential data processing offered algorithmic

Rajkumar et al. (2018) placed the modern approaches of machine learning in opposition to older expert systems by highlighting their data-driven nature: rule-based systems represented an explicit medical knowledge by creating an if-then rule by hand, whereas machine learning algorithms find statistical relationships and predictive relations in observational data. This shift in paradigm made AI systems be able to attack pattern recognition tasks, specifically in medical imaging, in which expert knowledge is hard to describe explicitly but there are many labeled examples available. The shift in learning to knowledge engineering to data-driven learning was a fundamental change in the course of medical AI research and application.



7.2 AI Applications in Diagnostic Medicine and Medical Imaging

7.2.1 Radiology and Diagnostic Imaging

Radiological interpretation is a canonical use of deep learning, and the original papers have shown that AI systems can perform at the level of a radiologist in various imaging modalities and clinical signs. The CheXNet created by Rajpurkar et al. (2017) is a 121-layer convolutional neural network that was trained using ChestX-ray14 dataset with 112,120 frontal-view chest X-rays and 14 pathological categories. In a validation study comparing CheXNet to practicing radiologists on pneumonia detection, the algorithm achieved an area under the receiver operating characteristic curve (AUROC) of 0.7632, exceeding the average radiologist performance of 0.7316. This study showed that deep learning could compete with expertise in a clinically relevant diagnostic task with a relatively widespread and available imaging modality.

Follow up studies generalized these results in terms of imaging modalities and anatomies. Ardila et al. (2019) designed a lung cancer screening end-to-end deep learning model based on low-dose computed tomography (CT) that was trained on the 42,290 screening CT scans. This model demonstrated an AUROC of 94.4 in the detection of malignant nodules, and a performance of 11 and 5 percentage points absolute reduction in false positive and false negative respectively, relative to unassisted radiologists, in a second validation cohort. It is notable that the system was capable of identifying cancers overlooked by radiologists during the initial clinical reads, which indicates that AI can be used as a second reader or quality assurance process.

In their study, McKinney et al. (2020) examined AI-based breast cancer screening with a deep learning trained on mammograms of more than 76,000 women in the United Kingdom and 15,000 women in the United States. The system performed better than individual radiologists with false positives being reduced by 5.7 percent and false negative being reduced by 9.4 percent in the UK dataset and similar results in the US validation cohort. The trial also showed that an AI system has the potential to preserve accuracy of screening replacing the second reader by the double-reading protocol in the UK, indicating that it might bring efficiency without affecting the quality of diagnostic findings. This discovery sparked significant debate on the best human-AI collaboration paradigms to support screening programs with some of the researchers supporting AI-enhanced steps and others warning about the needlessness to replace human skills (Freeman et al., 2021).

Chilamkurthy et al. (2018) made deep learning algorithms to detect and locate essential findings in head CT images with a sensitivity of over 90% to intracranial hemorrhage, mass effect, and midline shift with a processing time of less than one minute. Emergency department triage can potentially exceed in such systems because time-sensitive findings that need urgent attention can be identified immediately. This technology when commercialized under the Qure.ai platform can be considered one of the first examples of FDA cleared AI radiology systems to attain a real-world clinical implementation in numerous countries, such as India, United States, and Philippines.

Although these impressive performances were made, the critical inquiries have defied excitement with an appreciation of the methodological constraints and the difficulty in generalization. Nagendran et al. (2020) presented a systematic review of 20, 500 articles on AI applied in medical imaging, and only 516 articles provided external validation; moreover, most of them are highly prone to high risk of biasing associated with patient selection, reference standards, flow, and timing. The review found that the evidence on real-world performance and clinical utility was limited with limited studies assessing effect on the patient outcomes and not limited to the diagnostic accuracy. Park and Han (2018) also criticized the commonness of the so-called artificial intelligence chauvinism where hyping AI performance is done on the basis of ex-post facto studies based on curated datasets unrepresentative of clinical heterogeneity and the possibilities of diverse diagnostic challenges in practice.

Dataset shift and generalization failure are especially worrying processes in which AI models with high performance on training and validation datasets of particular institutions or populations would have a significant deterioration in performance on datasets of other sources. Zech et al. (2018) established that neural networks that were trained on the detection of pneumonia based on chest X-rays frequently acquired spurious associations with hospital-related imaging artifacts and scanner properties instead of meaningful pathological characteristics, leading to atrocious generalization to other datasets. The discovery highlights the significance of various training data, external validation, and continuous monitoring of performance in terms of AI systems implemented in heterogeneous clinical settings.

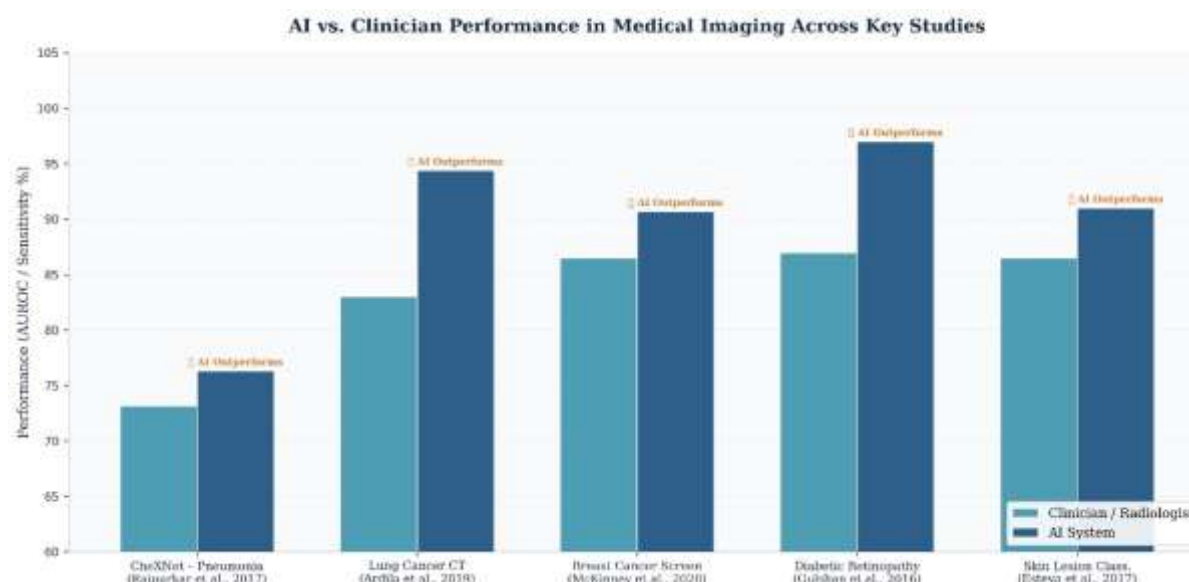


Figure 7.2: AI vs. Clinician Performance in Medical Imaging Across Key Landmark Studies.

7.2.2 Pathology and Histopathological Analysis

Digital pathology with deep learning has become the future of the enhancement of histopathological diagnosis, and can be used in cancer diagnosis, cancer grading and prediction of biomarkers prognostics. Esteva et al. (2017) established a deep convolutional neural network that was trained using 129,450 clinical images, which has been used to classify skin lesions with dermatologist-level accuracy (91% sensitivity and 90% specificity) in distinguishing between keratinocyte carcinomas and benign seborrheic keratoses and malignant melanomas and benign nevi. It employed transfer learning based on the use of a pre-trained network on general image classification tasks, which showed that AI systems do not necessarily have to be trained on medical images to achieve high performance.

By developing a multiple instance learning method that only needs slide-level labels, Campanella et al. (2019) overcame one of the core limitations of computational pathology the need to annotate the model exhaustively at the pixel level by pathologists, which necessitates slide-level annotations only. The system, trained on 12,160 gigapixel whole-slide images of prostate, basal cell carcinoma and breast cancer metastases, reported a cancer detection AUROC of more than 0.98. This methodological breakthrough greatly decreased the load of annotation, and thus, the large-scale training became possible and cost-effective.

Bera et al. (2019) analyzed AI applications in other areas of computational pathology rather than basic detection missions, such as molecular subtype prediction, genetic mutations, and treatment reactions in histopathological pictures. Research has shown that, using deep learning models, it is possible to predict the variables of microsatellite instability status, BRAF mutations, and KRAS mutations in colorectal cancer with routine histology, indicating that morphological patterns store genomic data that can be accessed by machine learning algorithms but not easily by human observers. Those capabilities beg interesting questions concerning whether AI machines detect authentic biological information or find small technical details in the preparation and digitization of images.

The workflow integration in the practice of AI in pathology faces different issues compared to radiology. Retamero et al. (2020) wrote about the complicated multi-step nature of pathology practice, such as gross examination, tissue processing, sectioning, staining, and microscopic interpretation, which AI tools have so far only tackled in limited sections of the workflow. In contrast to radiology where image interpretation is the main diagnostic output, the pathologist synthesizes morphological data with clinical history, ancillary analysis, and molecular data and has to interact with more complex information environments. Moreover, issues of standardization such as differences in tissue fixation, sectioning thickness and staining procedures across labs present further barriers to generalization of algorithms.

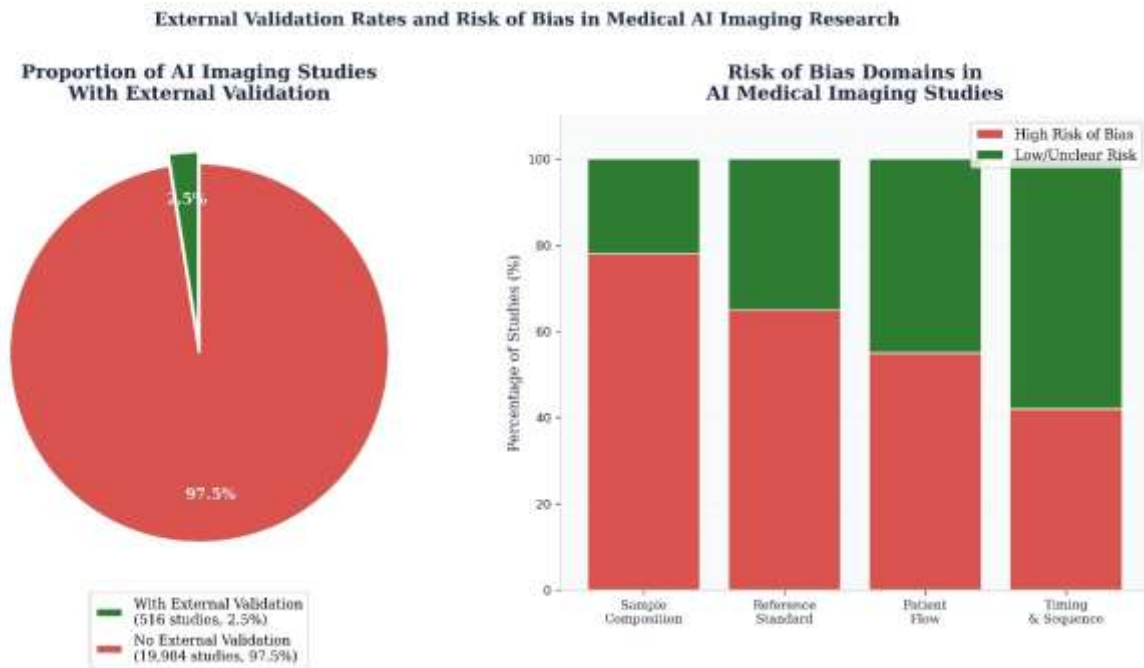


Figure 7.3: External Validation Rates and Risk of Bias Domains in AI Medical Imaging Studies.

7.2.3 Ophthalmology and Retinal Disease Detection

Ophthalmology has seen one of the most successful AI applications in a partnership between Google DeepMind and Moorfields Eye Hospital to create deep learning models to identify diabetic retinopathy and age-related macular degeneration in retinal fundus images and optical coherence tomography (OCT) images. Gulshan et al. (2016) created a deep neural network with diabetic retinopathy that was trained on 128,175 retinal images, which had a sensitivity of 90.3% and specificity of 98.1% when it came to referring diabetic retinopathy. Further testing in various clinical environments showed a sustained performance and the system obtained regulatory approvals such as FDA approval in 2018 and the CE mark in Europe.

De Fauw et al. (2018) also applied this to OCT imaging and created an AI system that identified more than 50 sight-threatening retinal conditions. The method used a two-step architecture, where a segmentation network was used to detect relevant anatomical structures and pathology, and then a classification network was used to produce referral recommendations and urgency. This design was more interpretable in that it gave visual explanations of what image regions were used to generate diagnostic outputs. In a prospective validation study of six retinal specialists assessing 997 OCT scans, the AI system produced performance comparable or superior to all human evaluators, and appropriate referral decisions in 94% cases.

The practice of AI-based diabetic retinopathy screening in the real-life context has provided important reports on the issues of implementing AI in real-life situations. Abramoff et al. (2018) outlined the clinical validation and approval procedure of the IDx-DR, an autonomous AI system that is planned to work without the supervision of physicians in the primary care facility. The landmark trial involved 900 subjects in primary care locations whereby the system had a sensitivity of 87.2% and a specificity of 90.7% as compared to grading of the Wisconsin Fundus Photograph Reading Center. It is important to note that FDA approval stated that the system did not require traditional physician interpretation to use- a important regulatory milestone since the first autonomous AI diagnostic system to have been cleared to clinical use in the United States.

Nevertheless, practical experience demonstrated that there are other aspects than diagnostic accuracy. In a study by Tufail et al. (2017), the outcomes of the implementation of an automated diabetic retinopathy screening system in the United Kingdom were reported and stated that 20-25% of the images were ungradable because of poor quality and needed to be reviewed by a human or a patient was to be recalled. Patient acceptance, integration with electronic health

records in workflow, liability aspect and reimbursement systems were also other barriers to implementation that had to be addressed in addition to designing the algorithm.

7.3 Predictive Analytics and Clinical Decision Support

In addition to diagnostic image analysis, artificial intelligence has been implemented in predictive modeling in structured electronic health care (EHR) data, clinical notes, lab findings and physiological surveillance to foretell clinical decline, readmissions, and response to treatment. These applications use the ability of machine learning to detect complex non-linear associations between high-dimensional clinical data that might not be detected by conventional statistical methods.

7.3.1 Early Warning Systems and Acute Care

A life-threatening organ dysfunction, due to an unregulated host response to infection, sepsis is the primary cause of mortality in hospitals and healthcare spending. The improvement of the results is possible by early recognition and intervention, which pushes the development of AI-based early warning systems. Nemat et al. (2018) trained a deep learning predictor on recurrent neural networks to predict sepsis based on physiological time series data, and found that their model has an area under the precision-recall curve of 0.31 to predict sepsis three hours before clinical identification of sepsis- a improvement that is impressively higher compared to the performance of other sepsis screening tools such as the Sequential Organ Failure Assessment (SOFA) score and Systemic Inflammatory Response Syndrome (SIRS) criteria.

Komorowski et al. (2018) used reinforcement learning, which is one of the machine learning paradigms in which the algorithms are trained to find the best decision policies when interacting with the environment to create treatment recommendations in the context of septic patients in the intensive care unit. The reinforcement learning model identified fluid resuscitation and vasopressor dosing strategies that were performance better in terms of survival, and the model-concordant treatments were less likely to result in mortality than the discordant treatments in the retrospective analysis, using data on 96,156 ICU stays. Although provocative, this article sparked a discussion of whether or not retrospective reinforcement learning is the right type of clinical decision support, since identified patterns of observation could be the results of confounding by indication and not a result of causal treatment effects (Saria, 2018).

Another widely researched application is hospital readmission prediction, which is due to the financial punishment as a part of the value-based care model, as well as the realization that avoidable readmission demonstrates the inadequacy of quality. In the study by Rajkomar et al. (2018), the authors created neural network models of predicting in-hospital mortality, 30-day unplanned readmission, and prolonged length of stay and discharge diagnoses based on the structured and unstructured EHR data of 216,221 hospitalizations in 2 health systems. The models recorded improved readmission and in-hospital mortality with AUROCs of 0.75 and 0.93, respectively as compared to the traditional models. Notably, the research used a general-purpose architecture that needs very little feature engineering, implying that the deep learning might decrease the tedious manual creation of predictive variables that are typical of conventional clinical risk scores.

Nevertheless, critical reviews have cast a doubt on the improvements of predictive accuracy to clinical utility. Goldstein et al. (2017) compared the Hospital Readmissions Reduction Program, indicating that even advanced prediction models have small levels of discrimination (AUROCs of 0.65-0.75) too small to be used to decide on individual level, but that they might be useful at the population level to allocate resources. As Billings et al. (2012) also claimed, the predictive models of readmission also have inherent shortcomings on the basis that readmissions are complex interactions among clinical, social, and behavioral factors that are not currently comprehensively reflected in medical data, and the stochastic error is significant and is not predictable. Such analyses highlight the significance of specifying correct use cases and performance limits instead of expecting the accuracy of prediction to be used as a measure of clinical significance.

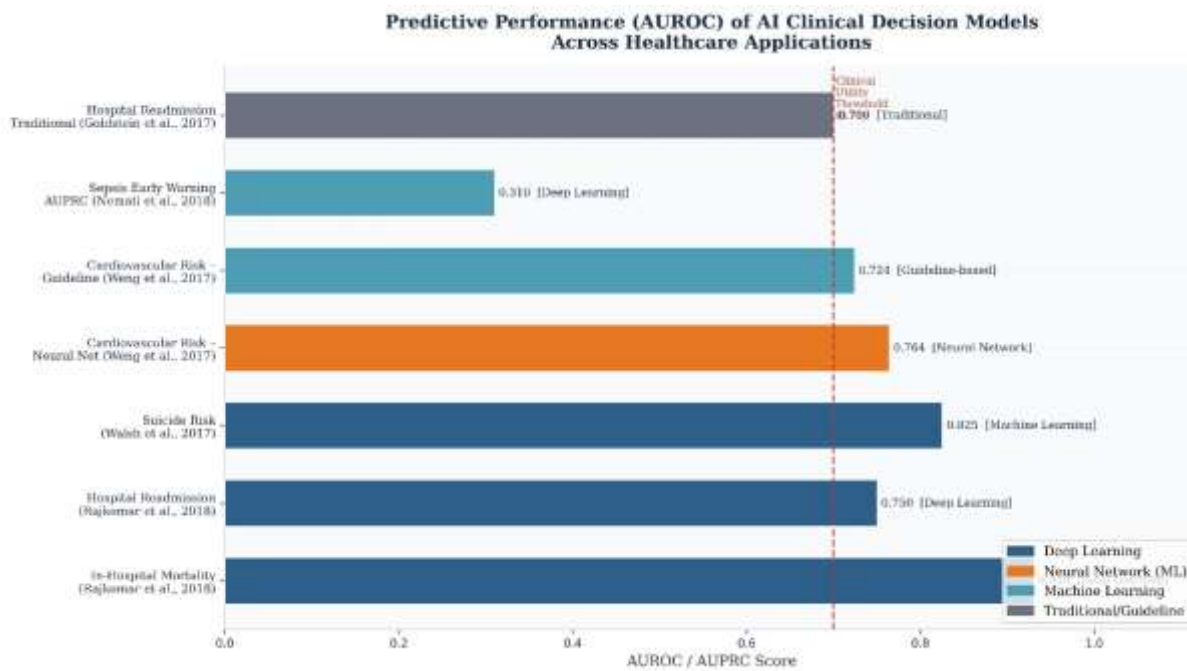


Figure 7.4: Predictive Performance (AUROC/AUPRC) of AI and Traditional Clinical Models.

7.3.2 Chronic Disease Management and Personalized Medicine

Applications of artificial intelligence in the management of chronic diseases are related to predicting disease progression, identifying high-risk patients to receive special treatment, and optimization of treatment plans. Reddy et al. (2016) created machine learning models that forecast personalized glycemic response to certain foods in diabetes management to allow personalized dietary advice. The participants recruited in the study gave dietary intake records and continuous glucose records in 800 people; whereby gradient boosting models gave significantly better prediction of postprandial glucose reactions compared to carbohydrate counting alone.

The fields of cancer prognosis and choice of treatment are some of the most active ones. Katzman et al. (2018) created DeepSurv, a deep learning-based survival analysis model that is superior to the Cox proportional hazards models in predicting patient outcomes and the interaction of treatment covariates with patient outcomes. The other category of AI use reviewed by Esteva et al. (2019) is related to the cancer imaging, pathology, genomics, and clinical data integration, with the focus on the possibility of multi-modal models consisting of various types of data to deliver more precise risk classification and treatment prescriptions than the single-modality models.

Machine learning methods have also been useful in cardiovascular disease prediction. Ambale-Venkatesh et al. (2017) used machine learning with cardiac magnetic resonance imaging data of the Multi-Ethnic Study of Atherosclerosis and constructed models that predict cardiovascular events and were better than the traditional risk scores, such as the Framingham Risk Score and pooled cohort equations. A comparison of four machine learning algorithms (random forest, logistic regression, gradient boosting, and neural networks) with the American College of Cardiology cardiovascular risk prediction guidelines using primary care records on 378,256 patients showed that neural networks had superior discrimination (AUROC 0.764) and detected 7.6% more patients who eventually developed cardiovascular events than guideline-based methods (Weng et al., 2017).

The application of pharmacogenomics and precision medicine is in the use of AI to predict individual responses to the drug depending on genetic, clinical, and demographic profiles. Nguyen et al. (2019) discussed the use of machine learning in the field of pharmacogenomics and identified successful applications in predicting adverse drug reactions, optimal dosing, and treatment efficacy in various clinical fields such as oncology, psychiatry, and cardiovascular medicine. Nevertheless, the review highlighted the recurring issues such as inadequacy of validation in populations,

sufficient attention to drug-drug interactions and polypharmacy in real-world settings, and regulatory ambiguity of clinical application of pharmacogenomic decision support instruments.

7.4 Natural Language Processing and Clinical Documentation

Clinical notes, such as physician notes, radiology notes, pathology notes, and discharge summaries, are rich data on issues of patient conditions, reason-making, and treatment plans, but are mostly inaccessible to computational analysis in their unstructured, textual form. Natural language processing (NLP) technologies allow the extraction, structuring, and analysis of information in clinical text and have found use in clinical decision support, quality measurement, pharmacovigilance, and clinical research.

Sheikhalishahi et al. (2019) presented a systematic review of the uses of NLP in nursing and clinical documentation, and found that uses of NLP in clinical documentation encompass finding symptoms, diagnoses, and treatments in clinical notes; detecting negative events and medication errors; and assessing documentation quality. The review observed that the accuracy of NLP was greatly different depending on the complexity of the task and characteristics of the text, although simple entity extraction had F-scores (harmonic mean of precision and recall) of over 0.90, whereas more complicated tasks, such as negation detection and time reasoning, were not easier.

The review of deep learning models of clinical NLP reported by Sheikhalishahi et al. (2019) has documented the shift toward the modern neural networks such as recurrent networks, convolutional networks, and attention-based transformer models as opposed to a traditional rule-based approach and classical machine learning. The review also noted BERT (Bidirectional Encoder Representations from Transformers) and its variants as especially successful, with models, which are pre-trained on large medical text corpora, performing state-of-the-art on tasks such as named entity recognition, relation extraction, and clinical text classification on task-specific datasets when fine-tuned.

Examples of the clinical potential of NLP are used. Murff et al. (2011) created NLP algorithms to learn the occurrence of post-operative complications based on clinical notes with a sensitivity of between 64-88% and positive predictive rates of between 72-97% for the different types of complications. With NLP FitzHenry et al. (2013) identified smoking status in clinical documentation, which is an important variable that is not always fully captured in structured EHR fields but got 93 percent accuracy. NLP helped Pivovarov and Elhadad (2015) detect patient subgroups with similar clinical presenting notes in the emergency department by finding clinically significant patient groups that were correlated with outcome and diagnosis.

The introduction of clinical language models has already gained traction due to the advances in general-domain NLP. Clinical BERT as developed by Alsentzer et al. (2019) was pre-trained on MIMIC-III clinical notes and showed better performance on clinical NLP tasks than on general-domain BERT. Lee et al. (2020) proposed BioBERT that was pre-trained with biomedical literature on PubMed and that demonstrated state-of-the-art performance on biomedical text mining applications such as named entity recognition, relation extraction, and question answering. Huang et al. (2019) created ClinicalBERT that is optimized to predict hospital readmission and diagnoses based on clinical notes, and it is superior to prior methods.

However, there are ongoing issues with clinical NLP. Clinical text is characterized by significant stylistic differences among providers, institutions and documentation contexts with abbreviations, medical terminology, incomplete sentences and grammatical abnormalities being a challenge to language models that are mainly trained on general well-formed text (Friedman et al., 1994). The temporal reasoning, knowing when events happened and how they are temporally related to each other, is a difficult concept in spite of clinical significance in discerning disease progression and response to treatment. Negation detection is used to tell affirmative results and those negated specifically (e.g. no evidence of pneumonia) or of family history, rather than that of the patient, with errors having a potential to cause false conclusions. Along with NLP uses, privacy is also an issue since mined data can accidentally disclose the confidential health data, so it needs to be de-identified and access-controlled.

7.5 AI in Drug Discovery and Development

The pharmaceutical research and development are traditionally characterized by the time of 10-15 years and costs that are above 2 billion dollars to introduce a new drug to the market, and the failure rates are particularly high at the stage of clinical trials (DiMasi et al., 2016). Applications of artificial intelligence in drug discovery include target

identification, compound screening, lead optimization, pharmacological property prediction and optimization of clinical trials, and its advocates believe that AI may significantly shorten the process of drug development and lower overheads and increase success chances.

7.5.1 Target Identification and Validation

The first step in drug discovery is to identify the biological targets, which are usually proteins and whose modification would help treat the disease. In their review of AI applications in target identification, King et al. (2009) characterize the usage of genomic, transcriptomic, proteomic, and phenotypic data to rank possible targets in the context of a machine learning model, based on anticipated disease-relevantness, drugability, and safety profiles. Protein-protein interaction networks Graph neural networks have been used to detect disease-related modules and therapeutic targets (Ding et al., 2020).

Nelson et al. (2019) used machine learning to combine various sources of data, such as genetic associations, gene expression, text mining of scientific literature and animal model data, to prioritize therapeutic targets in rheumatoid arthritis. The strategy was able to pinpoint known targets and suggest new ones which were later experimentally confirmed. This is an example on how AI can combine heterogeneous evidence that would be beyond the cognitive ability of human beings to integrate.

7.5.2 Compound Screening and Lead Optimization

Virtual screening uses computational science to screen large libraries of chemical compounds, making predictions about the most probable binding of the target protein by a compound, thus ranking compounds to be studied in experiments. Chen et al. (2018) performed a review of deep learning in drug discovery where neural networks are used to predict the molecular properties, binding affinities and toxicity profiles based on chemical structures alone. Such models are trained on encodings such as molecular fingerprints, graph encodings that encode chemical structure, and learned encodings that are learned by unsupervised pre-training on large datasets of compounds.

According to Stokes et al. (2020), the application of deep learning in antibiotic discovery was a landmark achievement wherein neural networks were trained to predict the existence of antibacterial activity of a compound on the basis of its molecular structure using data on 2,335 compounds. After screening 100 million molecules of the ZINC15 library (with an initial hit rate of 0.005), the model identified the compound halicin--a, previously tested drug for diabetes but not antibacterial activity- of interest- which in experimental validation showed potent antibacterial activity against drug-resistant organisms such as *Clostridioides difficile*, *Mycobacterium tuberculosis* and carbapenem-resistant *Enterobacteriaceae*. This publication showed that AI could find new compounds with new mechanisms of action that would otherwise remain hidden by the conventional medicinal chemistry methods that relied on structural analogs of already identified antibiotics.

The generation of new molecules is also a promising direction, with AI systems generating new molecules with particular properties as opposed to screening existing ones. Gomez-Bombarelli et al. (2018) trained variational autoencoders to design new molecules, to learn continuous representations of chemical space, which may be optimized to find new molecules by searching through the space with gradients. Zhavoronkov et al. (2019) used generative adversarial networks and reinforcement learning to design novel discoidin domain receptor 1 (DDR1) (a) kinase target inhibitors, creating 30,000 candidate molecules in 21 days and experimentally validating some of the potent and highly selective molecules. The lead compound entered into preclinical development showing a fully AI-enabled discovery pipeline that was significantly shorter than the standard one.

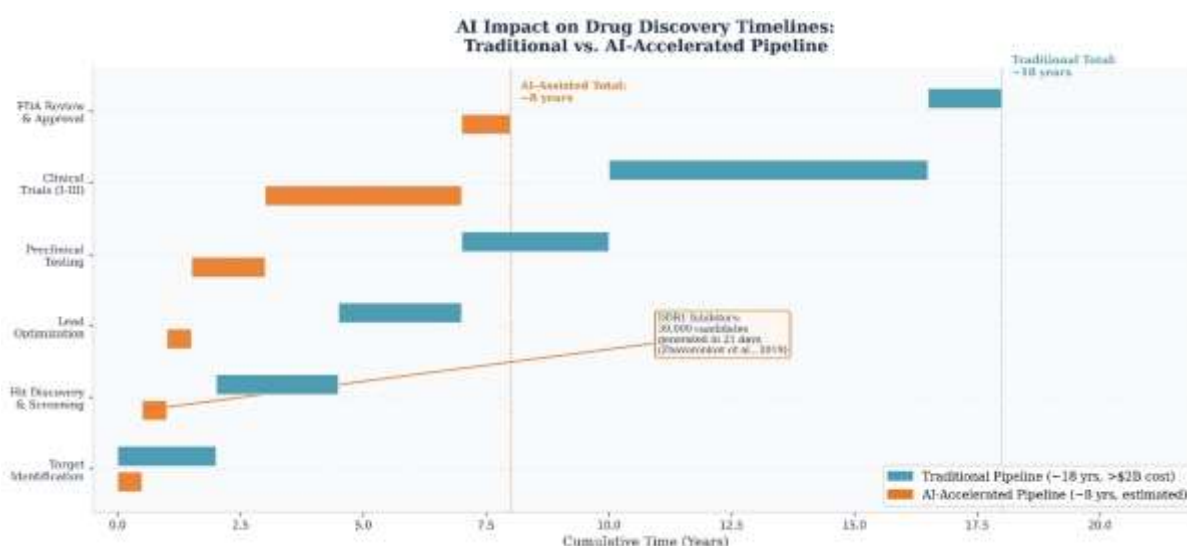


Figure 7.5: Comparison of Traditional vs. AI-Accelerated Drug Discovery Timelines.

7.5.3 Clinical Trial Optimization

The clinical trial uses of AI include optimization of patient recruitment, site selection, endpoint prediction, and protocol design. Harrer et al. (2019) have reviewed the approaches of machine learning in clinical trial design, indicating that methods have been used to predict patient enrollment rates, the most appropriate inclusion/exclusion criteria, and predict the trial outcome to facilitate adaptive trial design. Electronic health records can be analyzed through natural language processing to uncover potentially eligible patients, which can solve the problem of recruiting patients in clinical trials, becoming the source of trial delay and failure (Ni et al., 2015).

But critical attitudes put a damper on enthusiasm. Scannell et al. (2012) evaluated the waning efficiency of pharmaceutical R&D in the face of technological progress, suggesting that the productivity issue is inherently biological complexity, and difficult therapeutic targets as new therapy targets become harder, but regulatory obstacles, and not screening and optimization bottlenecks, which are the primary focus of AI. Bender and Cortes-Ciriano (2021) warned that present-day AI methods in drug discovery use the training data patterns associated with the correlations of the training data, which do not necessarily represent the causal biological mechanisms and potentially result in compounds that perform well based on the training data and fail to perform well in biological systems due to off-target effects, toxicity, or pharmacokinetic constraints of the training data.

This commercial environment is both highly confident and doubtful, with large pharmaceutical companies during this period like Pfizer, Novartis, and Sanofi partnering with AI companies like IBM Watson Health, and smaller biotechnology start-ups using AI as a core competency. Nonetheless, not many compounds developed by AI have made it to late-phase clinical trials, and the final determination of the effects of AI on drug discovery programs, budgets, and success rates is yet to be realized.

7.6 Robotics and Surgical AI

Another area of application is robotic surgical systems that are augmented with AI functionality and include platforms of robot-assisted surgery, surgical planning and navigation systems, and autonomous or semi-autonomous execution of surgical tasks. An example of first-generation surgical robotics that offer greater visualization and control under the supervision of a human surgeon instead of operating independently is the da Vinci Surgical System, which was produced by Intuitive Surgical and has been used in more than 8 million operations worldwide by 2020 (Intuitive Surgical, 2020).

Shademan et al. (2016) reported a supervised autonomous robotic surgery (STAR) system, which can operate soft tissue surgery, namely, intestinal anastomosis in ex vivo tissue, and in vivo porcine models. This system incorporating suturing, imaging, and near infrared fluorescent marks to monitor anatomy showed high-consistency in suture spacing and lower leak pressure than expert surgeons in ex vivo experiments, but took longer to complete the surgery. Although this work was more a research demonstration than a clinical deployment, it demonstrated technical capability of autonomous robotic execution of surgical skills.

Surgical uses of AI, further than robotic control, are preoperative planning, intraoperative decision support, and surgical skill evaluation. Vardazaryan et al. (2018) trained deep learning workflow recognition on video using surgical images to automatically recognize surgical phases and the use of surgical instruments with possible uses in surgical training, quality improvement, and intelligent operating rooms. Kitaguchi et al. (2020) used machine learning to evaluate surgical prowess using video recording of laparoscopic operations with an accuracy similar to human experts in the ability to differentiate between novice and experienced surgeons by the motion pattern of instruments.

The system of computer vision offers guidance through augmented reality to a surgeon during the procedure and superimposes anatomic structures, tumor limits, or intended resection edges on the clinical field of view (Bernhardt et al., 2017). The preoperative imaging to intraoperative anatomy registration should be accurate and take into account tissue deformation, breathing motion, and surgical manipulation, which can be solved with biomechanical modeling and AI-based deformable image registration methods.

Although proven with technology, there is mixed clinical evidence on the benefits of robotic or AI-enhanced surgery around the outcome of patients. There is an overall finding that systematic reviews have concluded that robotic surgery versus conventional laparoscopic methods produce comparable oncologic outcomes and complication rates, with increased operative time and significantly high costs, but possible shorter hospitalization and lessened blood loss (Sheetz et al., 2020). The high learning curves of robotic surgery, the high purchase cost (more than what a single system costs) and maintenance costs as well as disposable instrument costs make the subject of debate on the value proposition and proper usage of surgical robotics.

7.7 Mental Health and Behavioral Health Applications

Applications of AI in the mental health sector include chatbots, conversational agents in mental health therapy and counselling, predictive analytics in suicide risk and psychiatric crises, diagnostic aids in psychiatric conditions and digital phenotyping of mental states based on behavioral and physiological indicators using smartphone and wearable sensors.

7.7.1 Conversational AI and Digital Therapeutics

NLP-based chatbots offer affordable mental health services, which are mostly applicable to mild-to-moderate symptoms since conventional psychotherapy is inaccessible to a significant proportion of the population. Woebot is a text-based conversational agent that provides cognitive-behavioral therapy skills and was tested in a randomized controlled trial among anxiety and depressed college students (Fitzpatrick et al., 2017). The symptoms of depression and anxiety in the participants randomized to Woebot had significantly decreased in two weeks, relative to control participants offered an informational e-book, and the engagement of the participants and their satisfaction with using Woebot were high. The research established a feasibility, as well as initial effectiveness, but the long-term results and the comparison to the human-delivered therapy were not investigated.

A case in point of consumer interest in AI-mediated social interaction and emotional support, Replika is an emotionally supportive and chatbot intended for emotional support, not formal therapy, that acquired millions of users throughout the world (Skjuve et al., 2021). Nevertheless, there were also ethical concerns of proper limits of AI relations, risks of unhealthy dependency, and vulnerable people replacing chatbot interactions with the required professional mental health care or human relations.

Wysa and other mental health chatbots use AI-based conversations, along with evidence-based therapeutic strategies, mindfulness practices, and emotional monitoring. Inkster et al. (2018) administered a pragmatic trial to test Wysa, showing a substantial decrease in depressive symptoms in users who interacted with the chatbot but making it impossible to make causal inferences due to the lack of control. The strengths of chatbot interventions in terms of scalability and accessibility should be balanced with the implications of the absence of therapeutic depth, crisis management, and the risk of amplifying digital divide in case the AI tools will replace the traditional mental health services instead of complementing them.

7.7.2 Suicide Risk Prediction and Crisis Intervention

Trying to predict suicide risk is one of the difficult, but possibly highly effective AI applications because suicide causes more than 700,000 deaths worldwide every year as reported by the World Health Organization (2021). Conventional suicide risk assessment methods are based on clinical interviews and self-reported questionnaires with a small predictive validity. Walsh et al. (2017) used machine learning models based on electronic health record data of almost 6 million patients to predict suicide attempts with just under 0.80-0.85 values of AUROC, which is significantly higher than conventional screening devices. But such accurate values were still negative predictive values that were less than 1%, which doubted the clinical usefulness of such accurate values in terms of individual-level prediction, despite the potential of risk stratification on a population level.

Another strategy is the analysis of social media, and research has investigated linguistic patterns on posts before suicide attempts. Coppersmith et al. (2018) examined Twitter messages posted by people with known suicidal ideation, revealing linguistic features and both temporal relationship and temporal patterns involved in increased risks. Nevertheless, surveillance, consent, and the right intervention route are an ethical challenge that makes social media-based suicide prevention interventions challenging (Linthicum et al., 2019). What should platforms do when they have identified suicide risk based on algorithms-- direct clinical outreach, notifications to specified contacts, emergency services contact, and what accuracy levels would support the potentially stigmatizing measures?

7.8 Operational and Administrative AI Applications

Outside of direct clinical use, artificial intelligence can be used to alleviate healthcare administrative and operational heavy loads, which comprise nearly a quarter to a third of healthcare spending in the United States even amid other industrialized countries (Tseng et al., 2018). They can be used in the optimization of appointment scheduling, the management of patient flows, supply chain logistics, the detection of fraud, the administration of the revenue cycle, and the support of clinical documentation.

Bavafa et al. (2018) explored the use of machine learning to schedule appointments, formulating models that predict the likelihood of patient no-show to maximise the overbooking options according to the access / provider idle time and patient delays. The strategy brought about significant changes in comparison to the traditional scheduling regulations restricting provider idle time and decreasing the waiting period of the patients. Nonetheless, there were ethical issues of whether predictive overbooking on the basis of demographic and socioeconomic variables that relate to no-show probability may be considered a form of discrimination, where appointment opportunities are denied to the vulnerable groups that would most likely face the problem of transportation, childcare, or time constraints.

Another operational issue solved by using AI is the flow of patients and the allocation of resources in the emergency department. Raita et al. (2019) created machine learning models that predict disposition (admission and discharge) and length of stay in the case of the first triage data, allowing pre-emptive resource allocation and bed decision-making. These systems have the potential of resulting in fewer overcrowds and quicker services of patients who are in need of admission and avert hasty admissions of those whose conditions might clear with the short time observation.

NLP technologies help in the process of clinical documentation and this saves the administrative load that is a cause of physician burnout. Ambient clinical intelligence systems, including those created by Nuance Communications and Google, use speech recognition and NLP to generate clinical notes through conversations between the physician and the patient automatically and extract meaningful data and package it into standardized documentation formats (Neri et al., 2020). According to early reviews, such systems decrease documentation time and enhance physician satisfaction, but accuracy issues still exist, especially with complicated medical discussions, and concerns still exist about their effects on the dynamics of physician-patient communication when such discussions are electronically monitored and transcribed.

7.9 Ethical Considerations and Algorithmic Bias

A significant amount of academic literature has focused on the ethical aspects of AI in healthcare and studies have described the existence of algorithmic bias, issues of fairness, privacy considerations and general issues about the role AI should play in medical decision-making. This body of literature demonstrates the conflicts between the possible uses of AI in benefiting or furthering health inequalities.

7.9.1 Algorithmic Bias and Health Equity

Systematic errors that give unfair results that favor some groups of people over others have been reported in AI use in healthcare across all applications. The groundbreaking research by Obermeyer et al. (2019) investigated a highly popular commercial algorithm that predicts the healthcare costs to identify the patients who should be enrolled in the care management programs. It was found that even with equal risk scores, the Black patients were significantly iller in comparison with the White patients, and the algorithm was not systematically accurate at assessing the healthcare requirements of Black patients. The discrimination was due to the fact that the algorithm was trained to forecast healthcare expenses and not healthcare requirements and that historically black patients have been less taken care of by medical personnel as compared to white patients with the same ailments- something the algorithm learnt and enhanced its results. The bias that needs to be fixed would boost the proportion of Black patients under care management to 46.5 % rather than 17.7 % at a specified level of risk. The paper was an example to illustrate the ability of algorithms to incorporate and enforce existing inequalities by means of training them on biased historical data.

Gianfrancesca et al. (2018) addressed bias sources in AI healthcare applications, distinguishing between bias in training data (due to historical injustices) and bias in algorithms such as the inappropriate optimization objectives, and bias in application where the algorithms have interactions with human decision-makers, which can be biased and aversive to algorithms. The authors highlighted that bias reduction should not be limited to technical solutions but rather organizational promises to equity, multi-racial development teams, community involvement with the impacted communities, and repeated post-implementation monitoring on imbalanced effects.

Char et al. (2018) surveyed the topic of fairness in machine learning in healthcare and stated that various mathematical definitions of fairness are used, such as demographic parity (equality between selection rates of groups), equalized odds (equality between false positive and false negative rates), and predictive parity (equality between the positive predictive value of groups), which are mathematically incompatible, i.e. algorithms cannot realize all the fairness criteria at once when disease prevalence varies across groups. This impossibility conclusion suggests that fairness does not demand technical optimization, but value judgments as to the fairness criteria to be given priority to particular applications.

AI algorithms have also been reported to be biased in terms of gender. Larrazabal et al. (2020) reported that the performance of medical image classification algorithms was lower on female patients than on males both in various datasets and anatomies and is explained by the unbalanced training set where males were overrepresented and biological variations in disease presentation. Equally, performance differences based on age have been noted, whereby algorithm may not perform well with any pediatric or geriatric population when largely trained with adult data.

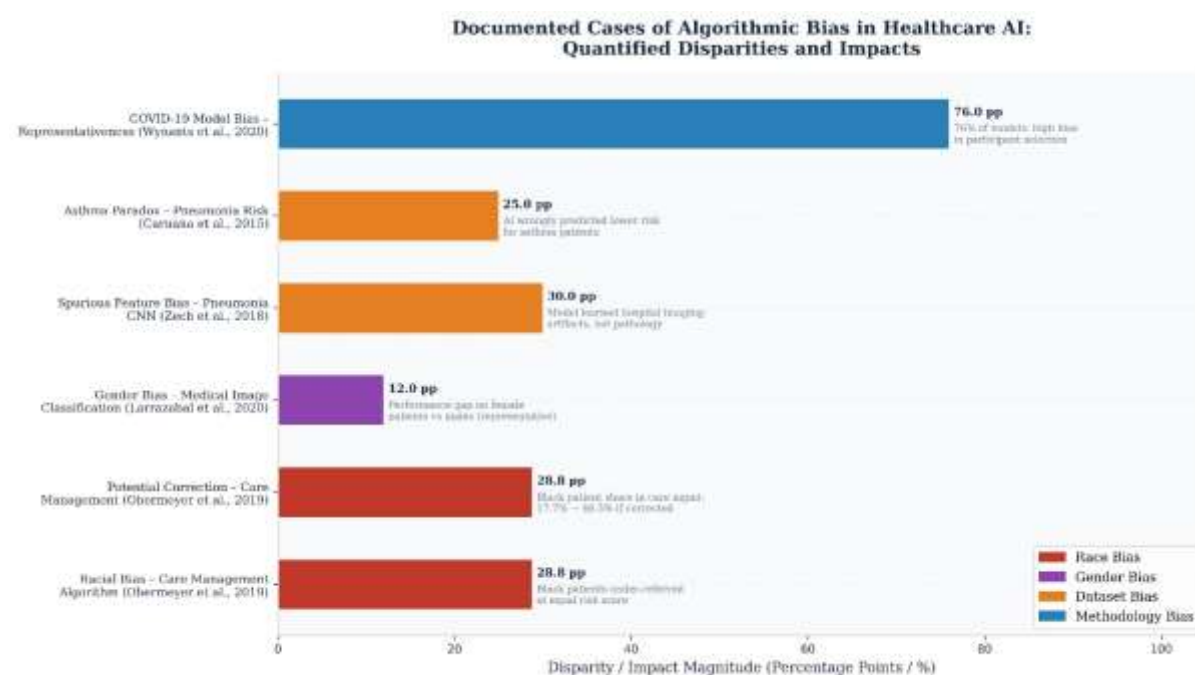


Figure 7.6: Documented Cases and Quantified Examples of Algorithmic Bias in Healthcare AI Systems.

7.9.2 Privacy, Consent, and Data Governance

The process of training AI systems needs access to huge volumes of sensitive medical data, which puts pressure on the collective good of better algorithms and individual rights to privacy. Price and Cohen (2019) considered privacy concerns in machine learning in healthcare, including that the standard privacy guarantees based on preventing re-identification of de-identified information are no longer adequate where algorithms can derive sensitive information about a data set using seemingly innocent data, or due to the unintended memorization of training data parameters. Membership inference attacks, in which attackers analyse whether the data of particular individuals were included in training sets, and model inversion attacks, in which attackers recreate training data based on model outputs, are new privacy attacks that are not well dealt with by existing regulatory frameworks.

In 2018, the General Data Protection Regulation (GDPR) of the European Union postulated very strict requirements such as a clear consent to process the data, a right to an explanation of the automated decisions, and the principle of minimum data. Such requirements present compliance difficulties in the creation of AI, especially in what is called the right to explanation of an algorithmic decision in case of complex deep learning models acting as black boxes (Goodman and Flaxman, 2017). Nevertheless, researchers argue as to whether or not technical explainability is consistent with meaningful transparency--detailed technical descriptions of how neural network computations are performed may be consistent with legal standards, but may also be unintelligible to patients and may not be actionable information as to why particular predictions were made or how to seek redress over the decision made.

Technical solutions, such as differential privacy, which introduces calibrated noise to the output of a model to mathematically quantify the privacy loss, are a potential solution, but tend to adversely affect model performance (Dwork and Roth, 2014). The pragmatic question starts to materialize: to which extent should accuracy be compromised in order to obtain stated privacy guarantees and by whom? Architectural alternatives to the protection of privacy through federated learning methods in which models develop on decentralized datasets without centrally aggregating raw data include federated learning methods, which have lower secrecy risks but entail potential technical challenges in convergence of models and performance (Rieke et al., 2020).

7.9.3 Transparency, Interpretability, and Explainability

The black box aspect of the complex machine learning models can be problematic to the medical situations where clinicians and the patients desire to know how to rationalize the diagnostic and treatment prescriptions that the models provide. Interpretability- the extent to which human beings are able to comprehend the causes of decisions- differs among AI methods with straightforward models such as decision trees providing transparency intrinsically and deep neural networks with millions of parameters unattainable to human understanding (Rudin, 2019).

There are several technical methods that deal with interpretability. Neural networks when used in attention pinpoint the input features that had the greatest impact on the outputs, which explains the post-hoc. Ribeiro et al. (2016) created LIME (Local Interpretable Model-Agnostic Explanations), an approximation of complex models in the local area with simpler ones that can be understood to explain single predictions. Saliency maps are visualization of the image regions whose influence in making a classification decision is the greatest in a computer vision application. Nevertheless, the research has shown that explanations are deceptive, indicating image regions that predict to correlate, yet do not causally predict, or that explanations that can be adversarially manipulated can be changed without causing prediction to change (Ghorbani et al., 2019).

Tonekaboni et al. (2019) have also doubted whether interpretability techniques can be used to enhance clinical trust and decision-making, and they have experimented on the topic by giving clinical predictions with and without explanations. Amazingly, explanations could sometimes lead to proper dependence on correct predictions and over-dependence on incorrect predictions when explanations were plausible even in respects of incorrect predictions-a warning point that superficial interpretability can raise falsely positive confidence and not real knowledge.

Rudin (2019) has presented compelling arguments to the effect that high-stakes areas such as healthcare must place greater emphasis on inherently interpretable models as opposed to complex black boxes that must be explained post-hoc, although this would involve a trade-off in accuracy. Interpretable models allow one to verify that an algorithm is making clinical recommendations instead of making spurious predictions and debug when these algorithms fail as well

as conform to the medical culture of focusing on understanding and not following statistical correlations. This view criticizes the tendencies of the field to put more emphasis on predictive accuracy over interpretability and whether marginal accuracy improvement by opaque models is worth losing transparency.

7.10 Regulatory Frameworks and Governance

The goal of healthcare AI regulation is to achieve safety and effectiveness, and create a stimulus to innovate, which is a difficult task because of the rapid change of technology, a variety of types of application, and global differences in regulatory philosophies. This part looks at regulatory strategies in key jurisdictions and policy discussions.

7.10.1 United States FDA Regulatory Framework

The United States Food and Drug Administration regulates medical AI as software as a medical device (SaMD) under the federal Food, Drug, and Cosmetic Act. Gerke et al. (2020) examined the changing practice of the FDA, stating that the previous method of regulating medical devices presupposed stable products with predetermined specifications, whereas AI systems can constantly learn and get better once deployed, which is known as an adaptive AI that does not fit into the traditional regulatory framework.

FDA has suggested a total product lifecycle regulatory model that differentiates between software precertification of software developers and between individual product authorization based on the fact that some organizations may always have high-quality and safe software whereas others would need more oversight (FDA, 2019). The agency came up with risk-based classification, which places stricter demands on AI systems that assist with the most important clinical decision-making than administrative applications with less risk. By 2021, more than 100 AI/ML-enabled medical devices, mostly in the field of radiology and cardiology, but not many of adaptive-based algorithms that adjust after deployment, were approved by the FDA (Benjamins et al., 2020).

Wu et al. (2021) examined the pattern of FDA authorization, where majority of the authorized AI medical devices were granted 510(k) clearance due to substantial equivalence to predicate devices, and not in de novo or premarket approval clearance that included more rigorous clinical evidence. The authors have raised the question of whether the equivalence-based clearance that is suitable in the case of conventional devices is relevant to the risks of AI-specific such as dataset shift, adversarial attack, and fairness. Moreover, the FDA does not have any specific requirements on algorithm transparency or bias evaluation, however, the draft guidance does take these factors into consideration (FDA, 2021).

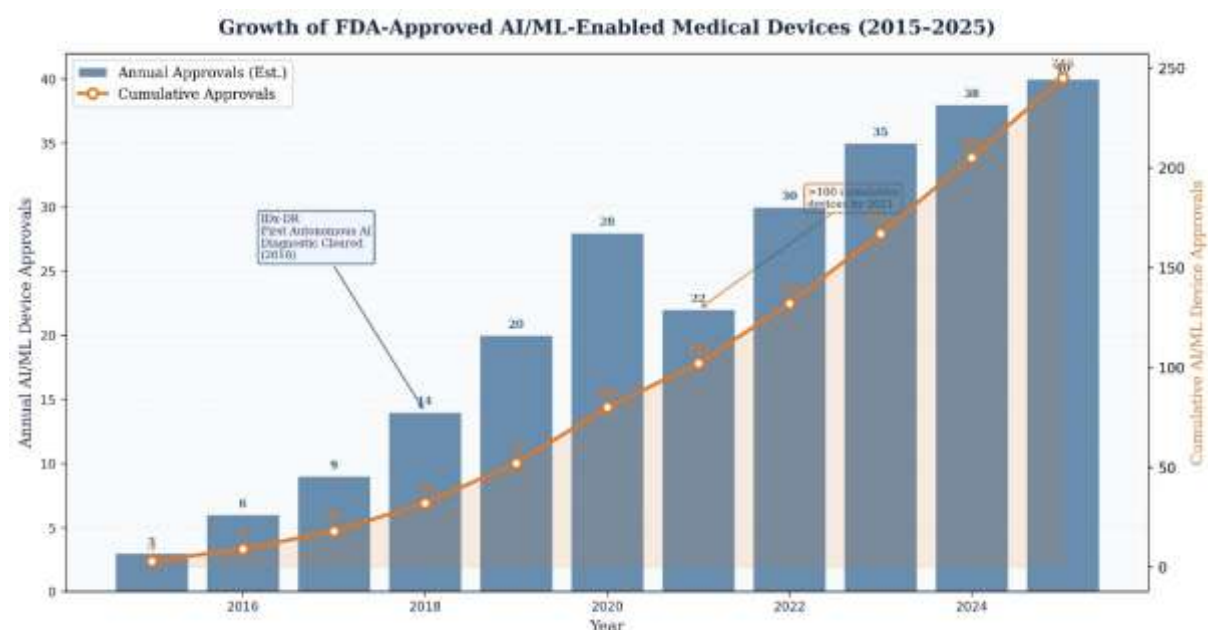


Figure 7.7: Growth of FDA-Approved AI/ML-Enabled Medical Devices (2015–2025).

7.10.2 European Union Regulatory Approach

Medical AI devices are regulated by the European Union via the Medical Device Regulation (MDR) and In Vitro Diagnostic Regulation (IVDR) passed in 2017 and 2022 respectively, as well as the proposed Artificial Intelligence Act (AIA) of horizontal regulation of AI in all spheres. The AIA is based on the principles of risk-based classification, where AI healthcare applications are classified as high-risk and demand conformity assessment, transparency requirements, and human supervision (European Commission, 2021).

Minssen et al. (2020) presented comparisons of the EU and the US regulatory systems, pointing out that in the former, the ex-ante third-party conformity assessment is done via the notified bodies, whereas the FDA performs the premarket review directly. The authors proposed that the EU solution can offer greater flexibility to the incremental innovations but could offer less standardization among the member states. In both jurisdictions, the issue of post-market surveillance and real-world performance monitoring-functionalities needed to identify performance deterioration, bias or safety concerns that may arise during clinical use-is problematic.

7.10.3 International Harmonization and Global Perspectives

International harmonization aims at creating uniformity in jurisdictions in order to access the global marketplace and eliminate overlapping regulatory costs. The International Medical Device Regulators Forum (IMDRF) came up with Software as a Medical Device framework that offer universal terminology and risk classification principles completed by regulators all over the world (IMDRF, 2017). Nevertheless, there are still significant differences in the aspects of evidence requirements, the approval timelines and the enforcement methods.

Low- and middle-income countries face a disproportional share of regulatory capacity issues and might not be equipped with knowledge and material to evaluate AI-specific medical devices. The regulatory considerations of AI healthcare technologies in resource-limited environments were discussed by Reddy et al. (2016), who highlighted the potential hazards related to implementing the systems that are not sufficiently validated in the locations with weak regulatory controls, which may cause harm to the vulnerable populations and undermine the trust in the positive effects of AI-based technologies. International standards, capacity-building efforts and regional regulation can help to alleviate such risks.

7.11 Economic and Commercial Dimensions

The healthcare sector of AI is characterized by a significant inflow of venture capital, market forecasts of swift growth, business model innovation, and economic studies of the cost-effectiveness and ROI. The knowledge of such economic facets is crucial in the case of commerce and healthcare administrators who have to make their technology adoption decisions.

7.11.1 Market Size, Growth, and Investment Trends

The global AI in healthcare market is projected to be worth in the range of 45-188 billion based on the definition and the methodology of the market, and the compound annual growth rates are estimated at 35-50% (Davenport and Kalakota, 2019). These forecasts are optimistic among investors and technology developers, but uncertainty on regulatory approvals, reimbursement policies, adoption rates by clinicians and competitive forces brings considerable variation.

In the late 2010s and early 2020s, venture capital investments in digital health and healthcare AI surged (e.g. the major funding rounds of Tempus (precision medicine data analytics), PathAI (computational pathology), Zebra Medical Vision (medical imaging AI), and Babylon Health (AI-powered primary care consultations). According to Rock Health, 378 deals and totaling 14.7 billion dollars in digital health funding in the United States in 2020 indicated that AI and machine learning represent some of the most popular investment themes (Rock Health, 2021).

Nevertheless, a number of high profile failures soothed blind admiration. IBM Watson Health, which was initially launched with great hype and collaborations with some of the largest cancer facilities to give treatment advice, has been criticized as lacking in clinical utility and eventually sold to a private equity firm in 2021 after years of lower performance than expected (Ross and Swetlitz, 2017; Strickland, 2019). Although not an AI company but more of a

laboratory testing company, Theranos was an excellent example of how the hype related to health technology innovation can exceed the scientific reality and defraud investors and patients, a cautionary measure that ought to encourage more critical questioning of healthcare AI-related claims.

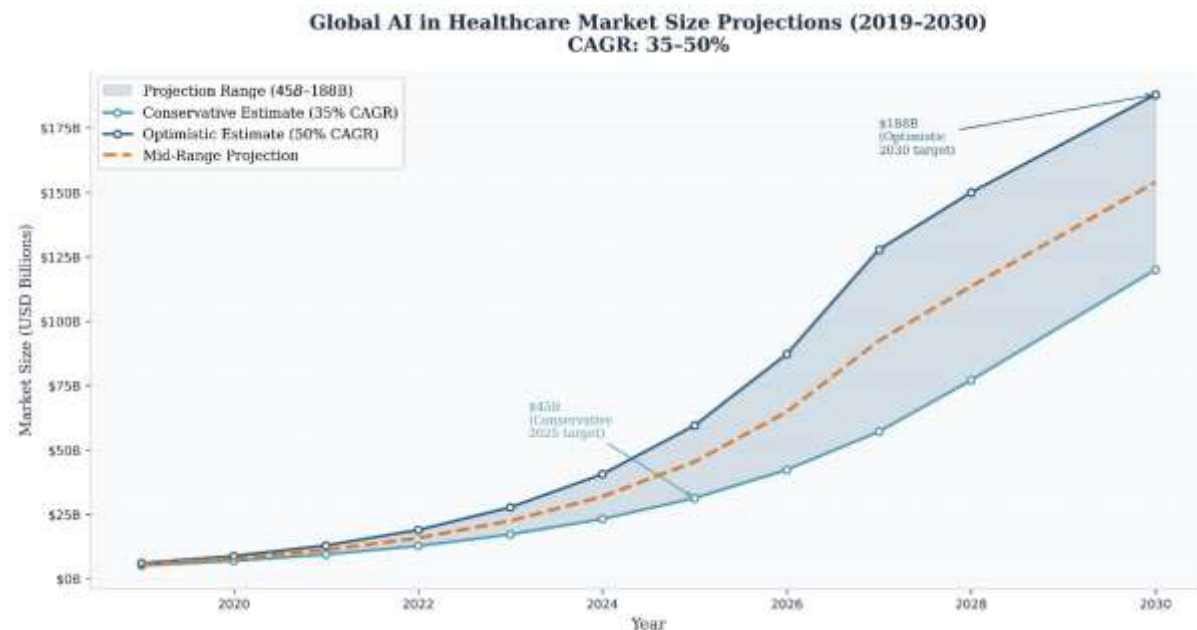


Figure 7.8: Global AI in Healthcare Market Size Projections (2019–2030).

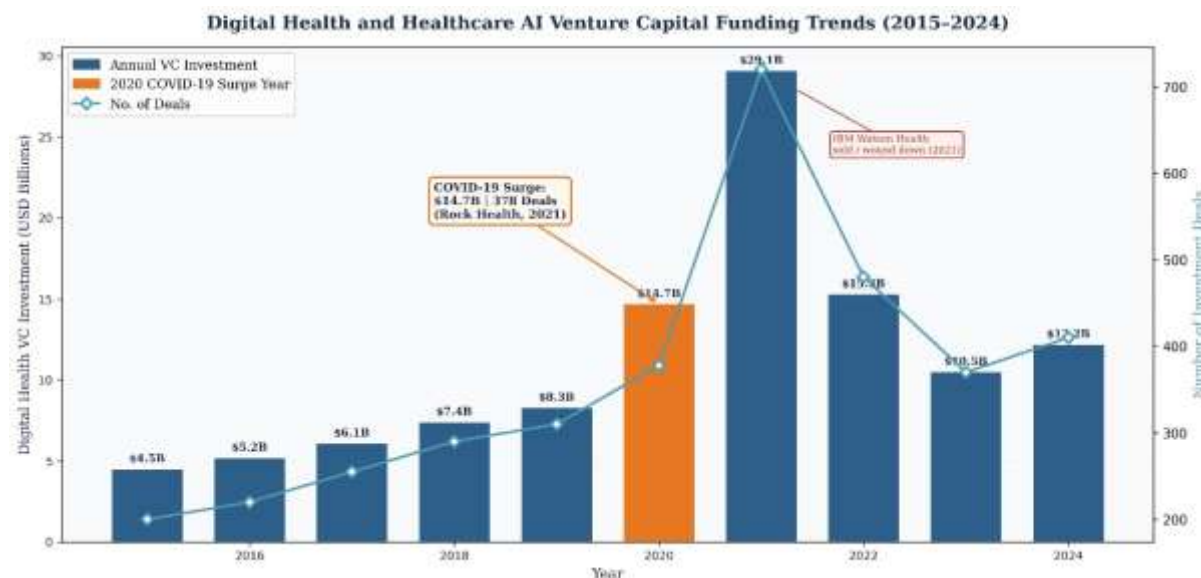


Figure 7.9: Digital Health and Healthcare AI Venture Capital Funding Trends (2015–2024).

7.11.2 Business Models and Value Propositions

Healthcare AI businesses use various business models such as software licensing to medical facilities, transaction fee per use, subscription access, data analytics, and direct-to-consumer applications. Hwang et al. (2019) undertook an overview of business models of clinical decision support AI, separating augmentation models in which AI aids human clinicians, and automation models in which AI aims to substitute human work. The authors hypothesized that augmentation models can have an easier regulatory route and professional adoption, whereas automation models have a higher potential to save costs, but will be resisted by healthcare workers and have liability issues.

The payment schemes play a crucial role in business feasibility. In the US, payers such as Medicare and commercial insurers started to develop Current Procedural Terminology (CPT) codes to cover particular AI applications- e.g. IDx-DR diabetic retinopathy screening was assigned a specific CPT code allowing billing. Nonetheless, the reimbursement practices are still fragmented and most of the AI services do not have a well-defined billing route posing a problem of

revenue unpredictability. Model value-based care, where reimbursement is not based on the quantity of services delivered but on the quality of services, theoretically encourages AI tools that reduce costs and enhance efficiency and quality, but in reality this type of model is not widespread (Blumenthal and McGinnis, 2015).

Direct-to-consumer AI health apps (such as symptom checkers, chatbots, wearable-built analytics) circumvent the old system of reimbursement by directly charging consumers. Nevertheless, it is associated with the issues of clinical validation of consumer health AI, the possibility of making wrong self-diagnosis and treatment, and the widening of health literacy gaps in case advanced technology becomes accessible to wealthy technologically advanced communities (Topol, 2015).

7.11.3. Cost-Effectiveness and Economic Impact

Economic analyses determine the value of AI technologies against the expenditure. Tran et al. (2019) performed a cost-effectiveness study of AI-based diabetic retinopathy screening, simulating the application in the United Kingdom in the National Health Service. The analysis established that AI screening was cost-effective relative to the existing methods of manual grading, with incremental cost-effectiveness ratios that are smaller than existing willingness-to-pay thresholds, and that this was primarily due to lower personnel costs and greater screening efficiency that allows broader program coverage.

Benjamins et al. (2020) warned that the issue of cost-effectiveness is determined by the context of implementation. AI systems that promise efficiency improvements can generate value only when health organizations are able to actually reduce staffing, reassign their staff to other duties, or accept more patients a change in their organization that goes beyond technology implementation. The fixed costs in AI implementation, such as data infrastructure, linking with current information systems, training, and maintenance may be higher than the savings of small healthcare organizations, possibly increasing the gap between better-resourced and resource-strained providers.

Healthcare AI has macroeconomic effects that encompass workforce implications, and the automation process may displace radiologists, pathologists, medical transcriptionists, and other professionals whose work is prone to being replaced by an algorithm. Nevertheless, other cases indicate that AI can complement instead of substitute employees, minimizing the workload of routine procedures and forming the need of new jobs in the management of AI systems, their support, and utilization (Brynjolfsson and McAfee, 2014). Their real employment effects will hinge on the technical capacity, regulatory choices, professional standoff or adjustments, and more macro labor market processes-issues, which cannot be addressed with longitudinal empirical research, which is not available.

7.12 Implementation Challenges and Sociotechnical Considerations

The difference between technological proofs in research and effective clinical performance is a crucial challenge among which the recent literature has abundant information. It is important to appreciate the sociotechnical factors, i.e. how technology interacts with human participants, organizational settings, and workflows in order to translate AI capabilities into actual clinical value.

7.12.1 Workflow Integration and Usability

The AI systems should also be able to fit within the current clinical processes instead of adding extra burdens. The barriers to implementation of clinical decision support systems have been studied by the researchers and poor system usability, non-integration with electronic health records, excessive notifications due to the alert fatigue, and disruption to workflow by system interactions were identified as the major obstacles. The same problems can be applied to AI-powered systems, which can produce recommendations that are to be reviewed by clinicians, documented, and acted upon accordingly.

Sendak et al. (2020) have outlined the operationalization of a sepsis early warning system by stating that technical model building was merely a small part of the implementation activity. The authors recorded the many efforts involved to engage stakeholders, redesign workflow, and set an alert threshold, develop training materials, and develop governance processes to measure performance and to respond to concerns, efforts not always considered in research publications that emphasize algorithmic innovation, rather than implementation science.

The design of the user interface has a major impact on adoption. An example provided by Caruana et al. (2015) is that a risk prediction model of pneumonia was unsuitably learning that patients with asthma incurred less risk - a spurious phenomenon that appeared due to the routine intensive care provided to asthmatic pneumonia patients, resulting in better outcomes. A purely correct black-box model would have advised non-provision of intensive care to asthmatic patients, due to this pattern, showing how interpretability allows identifying and mitigating dangerous errors before clinical catastrophe ensues.

7.12.2 Trust, Professional Acceptance, and Human-AI Collaboration

Doctors are divided in their attitude to AI as surveys demonstrate both enthusiasm and doubt. The literature review conducted by Blease et al. (2019) revealed that many clinicians were willing to use AI tools that might alleviate the burden of administration and increase the accuracy of the diagnosis; however, many were concerned about the potential loss of professional autonomy, deskilling, liability, and maintenance of the therapeutic relationship. The younger doctors and those closer to technology specialties were more accepting compared to older doctors and those in human-centric specialties such as psychiatry.

Attitude of the patients also differs. Nelson et al. (2020) discovered that the administrative and diagnostic image interpretation functions were perceived to be more comfortable by the patients with AI than complex treatment decisions and empathic communication, which are typically expected of human clinicians. Openness about the use of AI also affected positive responses, and patients indicated that they would prefer to be informed when AI was used in care consultation.

The idea of augmented intelligence, which places AI as a supplement, not as a substitution, of human judgment, has become popular as a conceptualization that minimises opposition and focuses on complementary capabilities of machine and human intelligence (Shortliffe and Sepulveda, 2018). Human beings are far better at contextual cognition, synthesis of psychosocial variables, emotional awareness, and moral judgment, whereas AI systems handle enormous volumes of data, detect minor patterns and exhibit uniformity. Probably the best set-ups would include cooperation, with each member taking advantage of their strengths.

Nonetheless, relying on AI advice too far is also a risk, especially when clinicians have automation bias, or a disproportionate level of trust in algorithmic results resulting in the inability to evaluate the information critically. Goddard et al. (2012) showed that this may happen in aviation and other fields where evidence has been found in medical case scenarios where providers sometimes accept false AI advice when that advice seems realistic especially when under time pressure or when there is cognitive overload.

7.12.3 Data Infrastructure and Interoperability

To implement AI, it is essential to have a strong data infrastructure such as electronic health record system with potential of collecting structured and unstructured information, data warehouses where the information is collated across various sources, interoperability standards that facilitate information sharing and data quality assurance measures. Kruse et al. (2014) discussed obstacles to health information exchange, and found that technical interoperability issues that emerge due to the use of different data standards, organizational barriers associated with competition issues regarding the exchange of information, and policy barriers such as privacy rules and lack of accountability on the quality of data.

The problem of data quality has a significant impact on the work of AI. Hersh et al. (2013) reported rampant issues of EHR data, such as missing values, incorrect entries, inconsistent coding habits, and imprecision in time. Machine learning models that are trained with low-quality data learn behaviors that are determined by data collection artifacts and not clinical reality, also known as a garbage in, garbage out effect that is especially problematic when machine learning models are applied to institutions whose data properties are unlike those in the training settings.

Findability, Accessibility, Interoperability, and Reusability of data are aspirational standards of healthcare data infrastructure that support AI (Wilkinson et al., 2016). Nonetheless, to obtain FAIR data, significant investment in standardization, metadata, persistent identifiers and governance is needed, which is competing with other healthcare priorities in resource-limited settings.

7.13 Global Perspectives and Applications in Resource-Limited Settings

Although much AI healthcare work and implementation has been done in high-income nations especially the United States, the United Kingdom and parts of Asia, there is an increasing interest in AI applications in low- and middle-income nations (LMICs) where healthcare demand is regularly highest and the shortage of health workforce is most severe.

Wahl et al. (2018) analyzed AI opportunities and challenges in global health by stating that AI may allow solving clinician shortages by assisting in diagnostic decision support, help manage supply chains in areas with inadequate medical facilities, and facilitate remote monitoring. Niche applications would be smartphone-based imaging to diagnose tuberculosis, malaria, and neglected tropical diseases, chatbots to provide simple health information in different languages, predictive models to diagnose disease outbreaks and allocate resources to response, etc.

GM implementation barriers in LMICs are however found to be significant. Cloud-based AI services are limited by limited internet connectivity and computational infrastructure, and local inference on smartphones or tablets can address this issue at least partially with the help of edge computing approaches (Farahani et al., 2020). Lack of data, especially labeled data to train supervised learning models, is a barrier as much of the conditions common in LMICs are underrepresented in the existing medical image databases or clinically acquired data to train models. High economic country datasets may not transfer learning in situations where the phenomenon of disease varies between populations because of genetic variation, co-morbidity, or late-Onset presentation typical of resource-constrained environments.

The issue of digital divide is whether AI medical technologies minimize or increase health disparities across the globe. Progressive views imagine democratization in which AI brings to remote areas the services of specialists of specialist level. Pessimistic scenarios predict the concentration of technologies in affluent areas with the infrastructure and resources necessary to deploy AI as poorer areas are under served- a trend that Hossain et al. (2019) described as algorithmic colonialism as technology created and used in affluent countries is inappropriate in the LMIC context but is pushed there anyway, potentially replacing locally relevant health solution.

7.14 COVID-19 Pandemic and Healthcare AI

The onset of the COVID -19 pandemic in late 2019 acted as a catalyst and a stress test to healthcare AI applications. Many AI systems were quickly created to screen, diagnose, prognose, repurpose and model epidemiology, and current AI systems in use, such as telemedicine and remote monitoring, were faster adopted due to the need to avoid in-person care following infection control measures.

In a systematic review of COVID-19 diagnosis and prognosis machine learning papers published as of June 2020, Roberts et al. (2021) found 232 articles that reported 299 models. Worryingly, the review concluded that all the models were not useable in the clinical setting with a high risk of bias with the frequent methodological problems of: training and testing on overlapping or poorly described datasets; non-external validation; unsuitable definition of outcomes; and limitations to reproducibility. Most researches have employed publicly available sets of images, of unknown provenance and quality, which may include duplicated images, or irrelevant information. The popularization of unproven models dangerously exposed users to clinical harm in the event of premature implementation, demonstrating the dangers of urgency and publication pressure undermining the validity of scientific behavior.

Similar results were obtained by Wynants et al. (2020) who found that the risk of bias was high in prediction models in 76% of the studies in the domain of participants (the majority because of non-representative sampling), and in 72% of the studies in the domain of outcomes. There were few studies that were detailed enough to be independently validated or replicated. The authors stressed that the low quality of model development that was witnessed in the course of COVID-19 was representative of systemic issues in clinical prediction model studies that are not limited to the pandemic, but the crisis just exacerbated the impacts of bad methodology.

During COVID-19, effective AI applications were in screening literature and extracting clinical insights using natural language processing, contact tracing applications (most of which were associated with privacy concerns and poor adoption), resource allocation optimization such as distributing ventilators and ICU beds (Vaishya et al., 2020). The implementation of telemedicine with artificial intelligence chatbots used to screen the effects of the symptom triggered

an excessive increase in scale, and healthcare organizations developed the capabilities of virtual care much faster than they would have under normal conditions.

The pandemic illuminated the need for strong AI frameworks that are resistant to the distribution changes- when data features vary with respect to training situations. COVID-19 presented new clinical manifestations, altered patient demographics with various population groups being more impacted than others over a period, and changes in healthcare practices that were characterized by more off-label prescription drugs and modified treatment regimens. In some situations, AI systems that were trained on the pre-pandemic data failed at the COVID-19 cases, which reflects the necessity to implement continuous monitoring and the possibility of updating the models under consideration (Heaven, 2019).

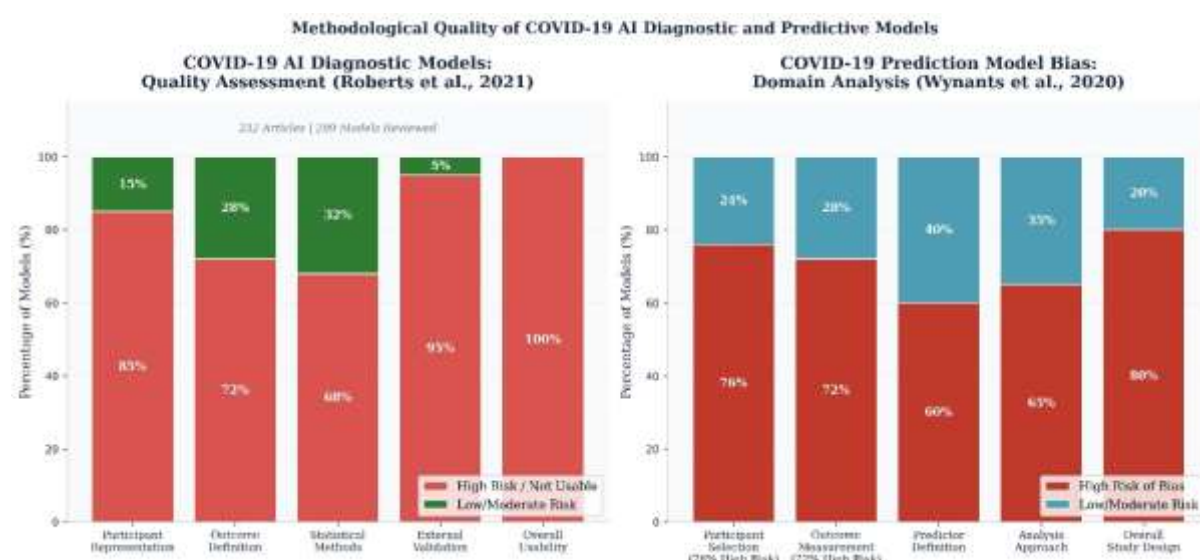


Figure 7.10: Methodological Quality and Risk of Bias in COVID-19 AI Diagnostic and Predictive Models.

7.15 Synthesis and Critical Analysis

The current literature demonstrates a discipline with impressive technological advances and unresolved issues in the areas of validation, generalization, ethics, regulation, and implementation. A number of cross-cutting themes come to the fore through domain synthesis.

First, it remains unaddressed that there is a gap between the impressive performance in the controlled research setting and impact in clinical reality. A variety of AI systems perform at the level of an expert in well-selected test sets but have significant challenges on heterogeneous clinical settings with diverse populations of patients, data quality, and workflows. This artificial intelligence chasm, which is similar to the translation gap between fundamental biomedical studies and clinical therapies, requires more focus on implementation science, human factors engineering and future assessment of clinical performance and cost-efficacy, but not just diagnostic accuracy.

Second, algorithmic bias and health equity issues are not sufficiently addressed as they are becoming more and more popular. The vast majority of AI systems are trained on the data of high-income populations, and high-income countries, which might not work well or lead to increased inequalities when used with low-represented ones. The response to bias must include technical remedies and institutional investing in equity, diverse teams of development, community involvement, and sustained surveillance, not always available in the existing practice.

Third, interpretability and trust are the persistent problems that lack solutions that are widely accepted. The debate around complex black-box models with higher predictive performance and more interpretable models with less complexity has not been resolved yet, and these trade-offs may be different depending on clinical usage. Laws and regulations do not specify interpretability requirements and decisions are yet to be made by individual developers and healthcare organizations.

Fourth, regulatory strategies are changing and falling behind technological competencies, especially in terms of continuously learning systems, bias assessment needs, and post-market surveillance. The efforts of harmonization at the

international level are encouraging but significant jurisdictional discrepancies are still present. The optimal degree of innovation promotion versus patient safeguarding is still a subject of debate where the dangers are too much regulation and the consequences of inadequate regulation may be a stifled innovation, which could by definition be beneficial and the absence of regulation that could allow the premature adoption of untested systems into service.

Fifth, economic and professional aspects demonstrate that there are conflicts between commercial and profit motives and the fact that healthcare is a social good. Although venture capital investment and market excitement have turned AI development into a fast-paced process, the sustainability of business models, their cost-efficiency, and fair access are raising concerns. The future of healthcare AI could be similar to the developments in pharmaceutical markets that have been characterized by the tendency to focus innovation on areas that are profitable and leave diseases with fewer resources and restricted environments under-innovated unless policy changes are implemented.

Lastly, the case of the COVID-19 pandemic served both as an exemplar of what AI could do, as well as a warning sign of the fragility of many systems that have been hastily deployed due to the urgency factor without much validation of their capabilities (or lack thereof). The experience highlights the necessity of the strong methodological norms, autonomous validation platforms, and fast but strict evaluation pathways between the speed of innovation and the rigor of science.

The reviewed literature confirms that artificial intelligence has a significant potential in changing the field of healthcare delivery, diagnosis, and treatment, and the evidence of this idea is shown in various fields. Nonetheless, to achieve this potential, it is necessary to deal with a complex set of issues that goes beyond the sphere of algorithmic development and includes issues of ethics, regulation, economics, and sociotechnical implementation. The field is at a peak junction wherein the basic research has already shown validity, regulatory systems are emerging, and clinical implementation is underway--but essential questions still emerge as to how to make AI a driving force towards rather than contributing to worse health outcome and disparities. The following segments of this dissertation define research gaps, objectives that would fill the gaps, as well as frameworks of further knowledge evolution of the role of AI in the future of healthcare.

8. Discussion, conclusion and recommendations.

8.1 Synthesis of Key Findings

The systematic review above shows a field at an extreme point of inflection. Across fifteen domains - spanning diagnostic AI, predictive analytics, drug discovery, surgical robotics, mental health applications, operational efficiency, and administrative automation - the literature indicates that AI in healthcare presents both extraordinary proof-concept successes in controlled environments and a set of entrenched structural, ethical, economic, and sociotechnical challenges to equitable and sustainable implementation. These two realities are not some abstract tensions; they have immediate and material implications on the investors, executives of health care, policymakers, and regulators who have to balance in this landscape in real time.

Regarding commerce and management, the most educative finding is that, not the technological aspects but mostly the sources of value and the sources of failure in healthcare AI are the most dominant. Performance on curated benchmarks through algorithms is required though not enough to be commercially viable. This dynamic is evidenced by the failure of the IBM Watson Health (Ross & Swetlitz, 2017; Strickland, 2019): a technologically advanced product was unable to go to commerce as it could not manage the organisational, clinical and reimbursement dynamics of actual healthcare systems. On the other hand, not only did IDx-DR have regulatory clearance, but also an acceptable level of accuracy in the algorithm, it also had an implemented reimbursement pathway and a feasible primary-care workflow (Abramoff et al., 2018). The takeaway of the lesson as an commerce student is clear: the use of technology is a needed but not a sufficient factor in commercial success in healthcare AI.

Technological Capabilities and Clinical Applications: The literature indicates that AI has impressive abilities in radiology (Rajpurkar et al., 2017; McKinney et al., 2020), predictive analytics (Rajkomar et al., 2018), natural language processing (Alsentzer et al., 2019), and drug discovery (Stokes et al., 2020). Nagendran et al. (2020) and Roberts et al. (2021) find, however, that most of the research studies available in the literature have high risk of bias, lack any external validation, and poorly defined outcome measures are the main characteristics. This disjunction between algorithmic

promise and clinical validity, in practice, has been the challenge of this generation of healthcare AI. This is management wise which requires due diligence frameworks of technology adoption that goes beyond vendor benchmark performance on the vendor report to independent validation of the outcome prospective data and disclosure of training data provenance as a prerequisite to procurement.

Real-World Effectiveness and Health System Impacts: Authentic patient-outcome benefit, rather than diagnostic-accuracy improvement, is relatively scanty. As Sendak et al. (2020) depict, the launch of a sepsis early-warning system necessitated a significant workflow redesign and stakeholder involvement and governance investment to a large extent surpassing the effort of developing the algorithmic implementation. Even statistically impressive readmission models, Goldstein et al. (2017) warn, have small clinical utility at an individual level. This lack of correlation between performance measures and health system value highlights the significance of implementation science and stringent health technology evaluation to commerce professionals analyzing AI investment, and overall, the value generated is often much less than publicized accuracy rates indicate.

Ethical Challenges, Bias, and Health Equity: Obermeyer et al. (2019) groundbreaking discovery, namely, that a popular commercial algorithm, when trained on historically inequitable utilisation data, systematically underestimated the healthcare needs of Black patients, is a case study in exemplifying how an algorithmic bias can further institutionalize and heighten existing health inequities. Gianfrancesca et al. (2018) and Char et al. (2018) also show that not a single mathematical definition of fairness fulfills all the equity criteria, such that fairness decisions require value judgements that cannot be assigned exclusively to the engineers. This poses both legal and reputational material risk to investors and corporate governance professionals: the use of algorithmic discrimination has become the subject of growing regulatory scrutiny, and organisations using biased systems are highly liable.

Regulatory Frameworks: Gerke et al. (2020), Minssen et al. (2020), and Wu et al. (2021) all present a regulatory landscape that is marked by a high rate of evolution, the existence of jurisdictional heterogeneity, and a number of remaining gaps-especially in the adaptive AI systems, post-market surveillance and mandatory bias assessment. Constructive but not comprehensive responses are the AI Act of the European Union and the total product lifecycle framework by the FDA. In the case of healthcare AI businesses, it is a serious commercial risk: organisations need to design both to meet the changing demands of multiple jurisdictions at the same time, and to feel blocked to spend capital and achieve long-term strategic planning when the approval process is unclear.

Commercial Dimensions, Business Models, and Market Dynamics: Rock Health (2021) reported that in 2020, the United States alone recorded investment of 14.7 billion in digital health, but market excitement should be constrained by the reimbursement reality (Hwang et al., 2019; Blumenthal and McGinnis, 2015). Most AI services do not have fixed billing codes, and revenue opportunities exist between the possibilities of technologies and the possibility to finance the enterprise. Davenport and Kalakota (2019) assume annual growth rates of 35-50, but based on those estimates are the guesses regarding the clearing rates of regulatory processes, the rate of clinical adoption, and the expansion of reimbursement that have a wide range of variation. Business models that are more focus on augmentation than automation seem to be more business-wise sustainable, as they tend to have fewer regulatory overheads and less resistance on the part of professionals, and are more apt to produce the evidence base necessary to achieve reimbursement recognition.

Implementation Problems and Organisational Barriers: It is a constant finding in the literature that sociotechnical factors, including workflow integration, clinician trust, data infrastructure quality, and organisational change capacity, are identified as the major determinants of implementation success or failure (Sendak et al., 2020; Blease et al., 2019; Kruse et al., 2014). Hersh et al. (2013) report the persistent data quality issues in electronic health record systems that deteriorate the performance of the model after deployment. The lesson to healthcare administrators is that successful AI implementation involves long-term organisational investment in data governance, change management, and user experience design costs that will consistently exceed the initial cost of technology acquisition and which must be factored into total cost of ownership calculations in the process of procurement assessment.

Research Gaps: Roberts et al. (2021) and Wynants et al. (2020) have revealed gaps in the methodology, specifically the lack of prospective external validation and reporting of fairness measures stratified by demographic subgroups, as the most significant research gaps in the current healthcare AI literature. Other gaps involve longitudinal cost-

effectiveness data in a varied payer setting, strict workforce influence research, and studies in low- and middle-income setting contexts (Wahl et al., 2018). In case of an interest in commerce-oriented research, the biggest gap is the lack of a serious health economic assessment of AI-enabled care models in comparison with the current practice in the entire variety of healthcare financing systems.

8.2 Implications to Commerce, Policy and Management

The existing healthcare AI investment opportunity can be characterized by features of the speculative timeline: influx of venture capital, optimistic market forecasts, and high-profile failures and real business success stories. The disappointing outcome of the IBM Watson Health unwind (Strickland, 2019) versus the long-running success of the da Vinci platform by Intuitive Surgical (Intuitive Surgical, 2020) and the long-term success of AI radiology companies represent the broad spectrum of outcomes that savvy investors need to calculate in healthcare AI portfolio values. Institutional investor portfolio construction may be beneficial in diversifying with diagnostic AI applications (less regulatory complexity, more established reimbursement pathways), operational and administrative AI (faster, more quantifiable ROI, less clinical risk), and early drug discovery AI (greater risk and return, longer payoff periods) which allows more risk-adjusted returns to optimise and benefits the overall improvement of healthcare systems.

In healthcare AI sustainability, reimbursement is of significant importance in comparison to the commercial aspect. Even products that have undergone a clinically proven and regulatory-approved development do not find it easy to turn a profit that would be worth their creation expenses. The transformative commercial effects of reimbursement recognition, as seen in the IDx-DR experience where Current Procedural Terminology code designation served as a straight ticket to sustainable revenue (Abramoff et al., 2018), prove the point. As a policy priority and a business approach, policy advocacy of outcome-linked reimbursement models, whereby payers will pay AI tools that can actually reduce downstream expenses or achieve an improved patient outcome. To develop an effective competitive advantage, healthcare AI companies must invest strategically in the field of health economic evidence generation because payers are increasingly demanding cost-effectiveness information as a prerequisite of coverage expansion (Blumenthal and McGinnis, 2015; Tran et al., 2019).

There is also a slow shift on the field, where pure software licensing is being replaced by integrated-service models including implementation support, performance monitoring, and guarantees of outcomes. Risk-sharing programs, in which AI vendors assume financial responsibility for clinical and operational outcomes, as opposed to simply licensing technology, are more compatible with commercial incentives and clinical worth and more likely to defeat institutional barriers to adoption. The healthcare executives considering AI vendors are encouraged to not just look at the performance of the algorithm and regulatory clearance, but also vendor ability to maintain an ongoing implementation relationship, data integration, and after-implementation performance reporting. Such outcome-dependent businesses are structurally encouraged by the move towards the environment of value-based care contracting (Blumenthal and McGinnis, 2015).

8.3 Practical Implementation Structure of Fair AI Implementation

Based on the synthesised evidence, this dissertation provides a five-pillar conceptual framework of equitable, commercially viable AI adoption in healthcare environments.

Pillar 1: Evidence-First Validation: Require prospective external validation in demographically diverse populations and the systematic use of reporting systems and demand performance data stratified by demographic subgroups as a requirement of regulatory clearance and institutional procurement. This solves the systemic bias issue presented by Nagendran et al. (2020) and Larrazabal et al. (2020), and gives the comparative effectiveness information needed by health technology assessment agencies to recommend reimbursement.

Pillar 2: Equity-by-Design: Conditional institutional clarity and institutional purchasing processes, which are formally required under the conditions of regulatory clearance and algorithmic bias audit as a formal requirement. Public release of training demographic data and the recorded bias reduction policies must be a compulsory measure so that buyers and regulators can make proper decisions regarding the suitability of deployment. The impossibility outcome reported by Char et al. (2018) implies that the definitions of fairness should not be covertly determined as their algorithmic formulations but be publicly chosen and announced.

Pillar 3: Sustainable Reimbursement Architecture: Design outcome-based reimbursement models with beneficial reward to AI tools that prove to enhance efficiency, quality or access. The natural market dynamics of concentration of investment in profitable markets would be offset by different reimbursement rates on applications used in underserved populations and under-resourced settings to provide commercial incentives that meet population health objectives (Hwang et al., 2019; Wahl et al., 2018).

Pillar 4: Sociotechnical Implementation Investment: Healthcare organisations implementing AI must invest equal resources in implementation infrastructure, including data governance, interoperability, workflow redesign, staff training, and continuous performance monitoring, understanding that the deployment environment, rather than the algorithm, dictates whether or not clinical value is actually achieved (Sendak et al., 2020; Wilkinson et al., 2016). The implementation quality standards should be integrated with the algorithmic performance requirements in regulatory and accreditation frameworks.

Pillar 5: Global Equity Commitments: Have tiered pricing, data sharing requirements, and capacity building requirements around AI implementations in low and middle income countries, counteract the risk of algorithmic colonialism via disproportionate benefits to communities whose data was used in model training (Hossain et al., 2019) through participating in the benefits fair and equally. Equity provisions should be included along with technical safety standards in international regulatory harmonisation via international regulatory bodies like the International Medical Device Regulators Forum (IMDRF, 2017).

8.4 Future Research Directions

There is a number of priority areas that should be addressed urgently by scholars. To begin with, future pragmatic clinical trials assessing the effects of AI on patient outcomes and healthcare costs not only in relation to diagnostic accuracy in curated test sets but also on diagnostic accuracy in general are urgently needed to get the evidence base beyond retrospective benchmark trials (Nagendran et al., 2020; Roberts et al., 2021). Second, a comparative effectiveness study comparing the various implementation models would present healthcare managers with evidence to action in the adoption of technologies. Third, research on the real employment effects of AI in medical professions longitudinally needs to be conducted to guide training programme and labour policy, beyond speculative forecasts (Brynjolfsson & McAfee, 2014). Fourth, health economic analyses performed in various payer contexts such as value-based care, single-payer systems, and under-resource settings would demystify the reimbursement approaches and make the decisions related to population coverage (Tran et al., 2019; Benjamins et al., 2020). Fifth, theoretically, to address demographic information concerning the disproportionate effects of AI, equity-oriented studies that directly assess the impact of AI on various demographic groups are required to support the transition to the practice (Obermeyer et al., 2019; Gianfrancesco et al., 2018). Lastly, privacy-preserving technologies like federated learning (Rieke et al., 2020) and differential privacy (Dwork and Roth, 2014) have to be evaluated systematically on their accuracy-privacy trade-offs to confirm their acceptability in the environment of various clinical and commercial scenarios.

8.5 Recommendations

To Policymakers: Speed up the creation of AI-specific regulatory frameworks that include compulsory bias assessment, standardised reporting of demographic performance, and harmonisation of international standards to decrease overlapping compliance costs (Gerke et al., 2020; Minssen et al., 2020). Create specific reimbursement programs on AI tools used in under-served specialties and locations. Establish independent, publicly available validation infrastructure the Cochrane Collaboration to clinical trials is comparable to provide performance assessments to vendors that are unbiased and reproducible and cannot be cherry-picked (FDA, 2021). Finance health AI literacy programmes that teach clinicians, patients, and administrators to be critical in regard to AI capabilities and limitations.

Granted to Healthcare Administrators: Implement a total-value-of-ownership model of AI purchasing that includes the implementation costs, data governance investment, employee training, change management, and continuous performance monitoring in addition to the acquisition costs. Obligate the vendors to report stratified equity, on a basis of demographic subgroups and post-deployment performance guarantees on contractual basis. Consider data governance and the interoperability infrastructure as necessary conditions before deploying AI and realise that the performance of an algorithm in the production context is limited by the quality of underlying data (Hersh et al., 2013; Wilkinson et al.,

2016). Create multidisciplinary AI governance boards - such as clinicians, ethicists, data scientists, patient representatives and legal counsel - to guide decisions on deployment and post-market performance monitoring.

To Investors: Focus on portfolio companies that have externally proven clinical outcomes data, known or reputable reimbursement structures, and clear equity reporting as well as have shown implementation relationships with healthcare organisations. Caution should be taken when it comes to any investment theses based on benchmark performance that is not independently validated. Participate in the governance of portfolio companies concerning disincentives to bias and regulatory oversight, understanding that algorithmic discrimination deems significant legal and reputational risks, which are not currently well-premeditated in early-stage valuation (Davenport and Kalakota, 2019; Rock Health, 2021). Evaluate the overall available market of AI applications based on the prospects of reimbursement rather than only clinical necessity.

To Regulators: Build flexible regulatory systems to regulate continuously learning AI systems, including pre-established performance limits on how they update, post-market surveillance with demographic disaggregation, and publicly available adverse events reporting, similar to the tasks of pharmacovigilance of pharmaceutical products (Wu et al., 2021). Make investments in global harmonisation of regulation in the framework of organisations like the IMDRF (2017) but retain the level of patient protection. Create relative avenues of AI applications to the context of low and middle-income countries that can consider the lack of resources without reducing safety (Reddy et al., 2016). Make public registries of deployed AI medical devices, including performance information broken down by deployment setting and patient demographic features.

8.6 Concluding Remarks

AI is the most impactful technological advancement in the healthcare sector since the formalisation of evidence-based medicine is not that its potential is fulfilled, but it is inexhaustible, and its consequences are complete. The evidence examined in this dissertation proves that AI will transform the diagnosis, therapeutics, drug development, healthcare operations, and the basic economics of healthcare delivery of all jurisdictions and care settings. It is no longer whether AI is going to revolutionize healthcare, but rather whether this revolution will be fair, commercially viable, and really helpful throughout the entire range of human health imperative.

This is a time when businesspeople and managerial practitioners should be entrepreneurial and humble in structure. The market potential is enormous--yet so are the organisational, regulatory, and ethical issues that define the success and failure of which business ventures. The timeline of healthcare AI is already filled with cautionary stories of over-engineered implementations, released without sufficient clinical validation, no reimbursement routes, no real implementation assistance and no equity aspects- each failure costing capital, trust and time to the actual achievement of true clinical worth. The businesses, which stand to create sustained business payoffs, would be those that see severe clinical validation, transparent equity accounting, extended reimbursement approach, and actual implementation collaboration not as compliance overheads but as the bases of competitive edge.

The general call, which cuts across all the stakeholder groups, is to make sure that the unprecedented analytical capability of artificial intelligence is used not just to lower costs to payers, create profits to investors, or decrease administrative overheads in clinicians, but in fact to make healthcare more accessible, higher-quality, and more equitable to everyone who needs it. To realise this potential, there must be continued and concerted effort on the part of the technologists, clinicians, economists, ethicists, policymakers, regulators and the communities of which they are a part. This dissertation has attempted to give conceptual underpinning and evidence base that commerce-oriented scholars and practitioners can play their part of significance and responsibility towards that common pursuit.

REFERENCES

- Abràmoff, M. D., Lavin, P. T., Birch, M., Shah, N., & Folk, J. C. (2018). Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. *NPJ digital medicine*, 1(1), 39.
- Alsentzer, E., Murphy, J., Boag, W., Weng, W. H., Jindi, D., Naumann, T., & McDermott, M. (2019, June). Publicly available clinical BERT embeddings. In *Proceedings of the 2nd clinical natural language processing workshop* (pp. 72-78).
- Ambale-Venkatesh, B., Yang, X., Wu, C. O., Liu, K., Hundley, W. G., McClelland, R., ... & Lima, J. A. (2017). Cardiovascular event prediction by machine learning: the multi-ethnic study of atherosclerosis. *Circulation research*, 121(9), 1092-1101.
- Ardila, D., Kiraly, A. P., Bharadwaj, S., Choi, B., Reicher, J. J., Peng, L., ... & Shetty, S. (2019). End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nature medicine*, 25(6), 954-961.
- Bavafa, H., Hitt, L. M., & Terwiesch, C. (2018). The impact of e-visits on visit frequencies and patient health: Evidence from primary care. *Management Science*, 64(12), 5461-5480.
- Bender, A., & Cortés-Ciriano, I. (2021). Artificial intelligence in drug discovery: what is realistic, what are illusions? Part 1: Ways to make an impact, and why we are not there yet. *Drug discovery today*, 26(2), 511-524.
- Benjamins, S., Dhunoo, P., & Meskó, B. (2020). The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. *NPJ digital medicine*, 3(1), 118.
- Bera, K., Schalper, K. A., Rimm, D. L., Velcheti, V., & Madabhushi, A. (2019). Artificial intelligence in digital pathology—new tools for diagnosis and precision oncology. *Nature reviews Clinical oncology*, 16(11), 703-715.
- Bernhardt, S., Nicolau, S. A., Soler, L., & Doignon, C. (2017). The status of augmented reality in laparoscopic surgery as of 2016. *Medical image analysis*, 37, 66-90.
- Billings, J., Blunt, I., Steventon, A., Georghiou, T., Lewis, G., & Bardsley, M. (2012). Development of a predictive model to identify inpatients at risk of re-admission within 30 days of discharge (PARR-30). *BMJ open*, 2(4), e001667.
- Blease, C., Kaptchuk, T. J., Bernstein, M. H., Mandl, K. D., Halamka, J. D., & DesRoches, C. M. (2019). Artificial intelligence and the future of primary care: exploratory qualitative study of UK general practitioners' views. *Journal of medical Internet research*, 21(3), e12802.
- Blumenthal, D., & McGinnis, J. M. (2015). Measuring vital signs: an IOM report on core metrics for health and health care progress. *Jama*, 313(19), 5-7.
- Brynjolfsson, E., & McAfee, A. (2014). *The second machine age: Work progress and prosperity in a time of brilliant technologies*. WW Norton & company.
- Buchanan, B. G., & Shortliffe, E. H. (1984). *Rule based expert systems: the mycin experiments of the stanford heuristic programming project (the Addison-Wesley series in artificial intelligence)*. Addison-Wesley Longman Publishing Co., Inc..
- Campanella, G., Hanna, M. G., Geneslaw, L., Mirafior, A., Werneck Krauss Silva, V., Busam, K. J., ... & Fuchs, T. J. (2019). Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nature medicine*, 25(8), 1301-1309.
- Caruana, R., Lou, Y., Gehrke, J., Koch, P., Sturm, M., & Elhadad, N. (2015, August). Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission. In *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1721-1730).
- Char, D. S., Shah, N. H., & Magnus, D. (2018). Implementing machine learning in health care—addressing ethical challenges. *The New England journal of medicine*, 378(11), 981.

- Chen, H., Engkvist, O., Wang, Y., Olivecrona, M., & Blaschke, T. (2018). The rise of deep learning in drug discovery. *Drug discovery today*, 23(6), 1241-1250.
- Chilamkurthy, S., Ghosh, R., Tanamala, S., Biviji, M., Campeau, N. G., Venugopal, V. K., ... & Warier, P. (2018). Deep learning algorithms for detection of critical findings in head CT scans: a retrospective study. *The Lancet*, 392(10162), 2388-2396.
- Coppersmith, G., Leary, R., Crutchley, P., & Fine, A. (2018). Natural language processing of social media as screening for suicide risk. *Biomedical informatics insights*, 10, 1178222618792860.
- Davenport, T., & Kalakota, R. (2019). The potential for artificial intelligence in healthcare. *Future healthcare journal*, 6(2), 94-98.
- De Fauw, J., Ledsam, J. R., Romera-Paredes, B., Nikolov, S., Tomasev, N., Blackwell, S., ... & Ronneberger, O. (2018). Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nature medicine*, 24(9), 1342-1350.
- DiMasi, J. A., Grabowski, H. G., & Hansen, R. W. (2016). Innovation in the pharmaceutical industry: new estimates of R&D costs. *Journal of health economics*, 47, 20-33.
- Wang, C., & Kurgan, L. (2019). Review and comparative assessment of similarity-based methods for prediction of drug-protein interactions in the druggable human proteome. *Briefings in bioinformatics*, 20(6), 2066-2087.
- Dwork, C., & Roth, A. (2014). The algorithmic foundations of differential privacy. *Foundations and trends® in theoretical computer science*, 9(3-4), 211-487.
- Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *nature*, 542(7639), 115-118.
- Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., ... & Dean, J. (2019). A guide to deep learning in healthcare. *Nature medicine*, 25(1), 24-29.
- European Commission. (2021). *Proposal for a regulation laying down harmonised rules on artificial intelligence*. COM(2021) 206 final.
- Farahani, B., Firouzi, F., & Chakrabarty, K. (2020). Healthcare iot. *Intelligent internet of things: from device to fog and cloud* (pp. 515-545). Cham: Springer International Publishing.
- FDA. (2019). *Proposed regulatory framework for modifications to artificial intelligence/machine learning (AI/ML)-based software as a medical device (SaMD)*. U.S. Food and Drug Administration.
- FDA. (2021). *Artificial intelligence and machine learning in software as a medical device*. U.S. Food and Drug Administration.
- FitzHenry, F., Murff, H. J., Matheny, M. E., Gentry, N., Fielstein, E. M., Brown, S. H., ... & Speroff, T. (2013). Exploring the frontier of electronic health record surveillance: the case of postoperative complications. *Medical care*, 51(6), 509-516.
- Fitzpatrick, K. K., Darcy, A., & Vierhile, M. (2017). Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. *JMIR mental health*, 4(2), e7785.
- Freeman, K., Geppert, J., Stinton, C., Todkill, D., Johnson, S., Clarke, A., & Taylor-Phillips, S. (2021). Use of artificial intelligence for image analysis in breast cancer screening programmes: systematic review of test accuracy. *bmj*, 374.
- Friedman, C., Alderson, P. O., Austin, J. H., Cimino, J. J., & Johnson, S. B. (1994). A general natural-language text processor for clinical radiology. *Journal of the American Medical Informatics Association*, 1(2), 161-174.
- Gerke, S., Minssen, T., & Cohen, G. (2020). Ethical and legal challenges of artificial intelligence-driven healthcare. *Artificial intelligence in healthcare* (pp. 295-336). Academic Press.

- Ghorbani, A., Abid, A., & Zou, J. (2019). Interpretation of neural networks is fragile. *Proceedings of the AAAI conference on artificial intelligence* (Vol. 33, No. 01, pp. 3681-3688).
- Gianfrancesco, M. A., Tamang, S., Yazdany, J., & Schmajuk, G. (2018). Potential biases in machine learning algorithms using electronic health record data. *JAMA internal medicine*, 178(11), 1544-1547.
- Gianfrancesco, M. A., Tamang, S., Yazdany, J., & Schmajuk, G. (2018). Potential biases in machine learning algorithms using electronic health record data. *JAMA internal medicine*, 178(11), 1544-1547.
- Goldstein, B. A., Navar, A. M., & Carter, R. E. (2017). Moving beyond regression techniques in cardiovascular risk prediction: applying machine learning to address analytic challenges. *European heart journal*, 38(23), 1805-1814.
- Gómez-Bombarelli, R., Wei, J. N., Duvenaud, D., Hernández-Lobato, J. M., Sánchez-Lengeling, B., Sheberla, D., ... & Aspuru-Guzik, A. (2018). Automatic chemical design using a data-driven continuous representation of molecules. *ACS central science*, 4(2), 268-276.
- Goodman, B., & Flaxman, S. (2017). European Union regulations on algorithmic decision-making and a “right to explanation”. *AI magazine*, 38(3), 50-57.
- Gulshan, V., Peng, L., Coram, M., Stumpe, M. C., Wu, D., Narayanaswamy, A., ... & Webster, D. R. (2016). Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *jama*, 316(22), 2402-2410.
- Harrer, S., Shah, P., Antony, B., & Hu, J. (2019). Artificial intelligence for clinical trial design. *Trends in pharmacological sciences*, 40(8), 577-591.
- Heaven, D. (2019). Why deep-learning AIs are so easy to fool. *Nature*, 574(7777), 163-166.
- Hersh, W. R., Weiner, M. G., Embi, P. J., Logan, J. R., Payne, P. R., Bernstam, E. V., ... & Saltz, J. H. (2013). Caveats for the use of operational electronic health record data in comparative effectiveness research. *Medical care*, 51, S30-S37.
- Hossain, M. S., Muhammad, G., & Guizani, N. (2020). Explainable AI and mass surveillance system-based healthcare framework to combat COVID-19 like pandemics. *IEEE network*, 34(4), 126-132.
- Huang, K., Altosaar, J., & Ranganath, R. (2019). Clinicalbert: Modeling clinical notes and predicting hospital readmission. *arXiv preprint arXiv:1904.05342*.
- Hwang, T. J., Kesselheim, A. S., & Vokinger, K. N. (2019). Lifecycle regulation of artificial intelligence—and machine learning—based software devices in medicine. *Jama*, 322(23), 2285-2286.
- IMDRF. (2017). *Software as a Medical Device (SaMD): Clinical evaluation*. International Medical Device Regulators Forum.
- Inkster, B., Sarda, S., & Subramanian, V. (2018). An empathy-driven, conversational artificial intelligence agent (Wysa) for digital mental well-being: real-world data evaluation mixed-methods study. *JMIR mHealth and uHealth*, 6(11), e12106.
- Intuitive Surgical. (2020). *Annual report 2020*. Intuitive Surgical, Inc.
- Katzman, J. L., Shaham, U., Cloninger, A., Bates, J., Jiang, T., & Kluger, Y. (2018). DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network. *BMC medical research methodology*, 18(1), 24.
- King, R. D., Rowland, J., Oliver, S. G., Young, M., Aubrey, W., Byrne, E., ... & Clare, A. (2009). The automation of science. *Science*, 324(5923), 85-89.
- Kitaguchi, D., Takeshita, N., Matsuzaki, H., Oda, T., Watanabe, M., Mori, K., ... & Ito, M. (2020). Automated laparoscopic colorectal surgery workflow recognition using artificial intelligence: experimental research. *International journal of surgery*, 79, 88-94.

- Komorowski, M., Celi, L. A., Badawi, O., Gordon, A. C., & Faisal, A. A. (2018). The artificial intelligence clinician learns optimal treatment strategies for sepsis in intensive care. *Nature medicine*, 24(11), 1716-1720.
- Kruse, C. S., Regier, V., & Rheinboldt, K. T. (2014). Barriers over time to full implementation of health information exchange in the United States. *JMIR Medical informatics*, 2(2), e3625.
- Larrazabal, A. J., Nieto, N., Peterson, V., Milone, D. H., & Ferrante, E. (2020). Gender imbalance in medical imaging datasets produces biased classifiers for computer-aided diagnosis. *Proceedings of the National Academy of Sciences*, 117(23), 12592-12594.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, 521(7553), 436-444.
- Lee, J., Yoon, W., Kim, S., Kim, D., Kim, S., So, C. H., & Kang, J. (2020). BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*, 36(4), 1234-1240.
- Linthicum, K. P., Schafer, K. M., & Ribeiro, J. D. (2019). Machine learning in suicide science: Applications and ethics. *Behavioral sciences & the law*, 37(3), 214-222.
- McKinney, S. M., Sieniek, M., Godbole, V., Godwin, J., Antropova, N., Ashrafi, H., ... & Shetty, S. (2020). International evaluation of an AI system for breast cancer screening. *Nature*, 577(7788), 89-94.
- Miller, R. A. (1994). Medical diagnostic decision support systems—past, present, and future: a threaded bibliography and brief commentary. *Journal of the American Medical Informatics Association*, 1(1), 8-27.
- Minssen, T., Gerke, S., Aboy, M., Price, N., & Cohen, G. (2020). Regulatory responses to medical machine learning. *Journal of Law and the Biosciences*, 7(1), 1-10.
- Murff, H. J., FitzHenry, F., Matheny, M. E., Gentry, N., Kotter, K. L., Crimin, K., ... & Speroff, T. (2011). Automated identification of postoperative complications within an electronic medical record using natural language processing. *Jama*, 306(8), 848-855.
- Nagendran, M., Chen, Y., Lovejoy, C. A., Gordon, A. C., Komorowski, M., Harvey, H., ... & Maruthappu, M. (2020). Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies. *bmj*, 368.
- Nelson, C. A., Butte, A. J., & Baranzini, S. E. (2019). Integrating biomedical research and electronic health records to create knowledge-based biologically meaningful machine-readable embeddings. *Nature communications*, 10(1), 3045.
- Nelson, C. A., Pérez-Chada, L. M., Creadore, A., Li, S. J., Lo, K., Manjaly, P., ... & Mostaghimi, A. (2020). Patient perspectives on the use of artificial intelligence for skin cancer screening: a qualitative study. *JAMA dermatology*, 156(5), 501-512.
- Nemati, S., Holder, A., Razmi, F., Stanley, M. D., Clifford, G. D., & Buchman, T. G. (2018). An interpretable machine learning model for accurate prediction of sepsis in the ICU. *Critical care medicine*, 46(4), 547-553.
- Neri, E., Coppola, F., Miele, V., Bibbolino, C., & Grassi, R. (2020). Artificial intelligence: Who is responsible for the diagnosis?. *La radiologia medica*, 125(6), 517-521.
- Nguyen, P. A., Born, D. A., Deaton, A. M., Nioi, P., & Ward, L. D. (2019). Phenotypes associated with genes encoding drug targets are predictive of clinical trial side effects. *Nature communications*, 10(1), 1579.
- Ni, Y., Wright, J., Perentesis, J., Lingren, T., Deleger, L., Kaiser, M., ... & Solti, I. (2015). Increasing the efficiency of trial-patient matching: automated clinical trial eligibility pre-screening for pediatric oncology patients. *BMC medical informatics and decision making*, 15(1), 28.
- Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464), 447-453.
- Park, S. H., & Han, K. (2018). Methodologic guide for evaluating clinical performance and effect of artificial intelligence technology for medical diagnosis and prediction. *Radiology*, 286(3), 800-809.

- Pivovarov, R., & Elhadad, N. (2015). Automated methods for the summarization of electronic health records. *Journal of the American Medical Informatics Association*, 22(5), 938-947.
- Price, W. N., & Cohen, I. G. (2019). Privacy in the age of medical big data. *Nature medicine*, 25(1), 37-43.
- Raita, Y., Goto, T., Faridi, M. K., Brown, D. F., Camargo Jr, C. A., & Hasegawa, K. (2019). Emergency department triage prediction of clinical outcomes using machine learning models. *Critical care*, 23(1), 64.
- Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. *New England Journal of Medicine*, 380(14), 1347-1358.
- Rajkomar, A., Oren, E., Chen, K., Dai, A. M., Hajaj, N., Hardt, M., ... & Dean, J. (2018). Scalable and accurate deep learning with electronic health records. *NPJ digital medicine*, 1(1), 18.
- Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., ... & Ng, A. Y. (2017). Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. *arXiv preprint arXiv:1711.05225*.
- Reddy, M., Pesl, P., Xenou, M., Toumazou, C., Johnston, D., Georgiou, P., ... & Oliver, N. (2016). Clinical safety and feasibility of the advanced bolus calculator for type 1 diabetes based on case-based reasoning: a 6-week nonrandomized single-arm pilot study. *Diabetes technology & therapeutics*, 18(8), 487-493.
- Retamero, J. A., Aneiros-Fernandez, J., & Del Moral, R. G. (2020). Complete digital pathology for routine histopathology diagnosis in a multicenter hospital network. *Archives of pathology & laboratory medicine*, 144(2), 221-228.
- Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should i trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1135-1144).
- Rieke, N., Hancox, J., Li, W., Milletari, F., Roth, H. R., Albarqouni, S., ... & Cardoso, M. J. (2020). The future of digital health with federated learning. *NPJ digital medicine*, 3(1), 119.
- Roberts, M., Driggs, D., Thorpe, M., Gilbey, J., Yeung, M., Ursprung, S., ... & Schönlieb, C. B. (2021). Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans. *Nature Machine Intelligence*, 3(3), 199-217.
- Rock Health. (2021). *2020 year end digital health funding report*. Rock Health.
- Ross, C., & Swetlitz, I. (2017). IBM pitched its Watson supercomputer as a revolution in cancer care. It's nowhere close. *Stat*, 5.
- Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature machine intelligence*, 1(5), 206-215.
- Saria, S. (2018). Individualized sepsis treatment using reinforcement learning. *Nature medicine*, 24(11), 1641-1642.
- Scannell, J. W., Blanckley, A., Boldon, H., & Warrington, B. (2012). Diagnosing the decline in pharmaceutical R&D efficiency. *Nature reviews Drug discovery*, 11(3), 191-200.
- Sendak, M. P., Gao, M., Brajer, N., & Balu, S. (2020). Presenting machine learning model information to clinical end users with model facts labels. *NPJ digital medicine*, 3(1), 41.
- Shademan, A., Decker, R. S., Opfermann, J. D., Leonard, S., Krieger, A., & Kim, P. C. (2016). Supervised autonomous robotic soft tissue surgery. *Science translational medicine*, 8(337), 337ra64-337ra64.
- Sheetz, K. H., Claflin, J., & Dimick, J. B. (2020). Trends in the adoption of robotic surgery for common surgical procedures. *JAMA network open*, 3(1), e1918911.
- Sheikhalishahi, S., Miotto, R., Dudley, J. T., Lavelli, A., Rinaldi, F., & Osmani, V. (2019). Natural language processing of clinical notes on chronic diseases: systematic review. *JMIR medical informatics*, 7(2), e12239.

- Shortliffe, E. (Ed.). (2012). *Computer-based medical consultations: MYCIN* (Vol. 2). Elsevier.
- Shortliffe, E. H., & Sepúlveda, M. J. (2018). Clinical decision support in the era of artificial intelligence. *Jama*, 320(21), 2199-2200.
- Skjuve, M., Følstad, A., Fostervold, K. I., & Brandtzaeg, P. B. (2021). My chatbot companion-a study of human-chatbot relationships. *International Journal of Human-Computer Studies*, 149, 102601.
- Stokes, J. M., Yang, K., Swanson, K., Jin, W., Cubillos-Ruiz, A., Donghia, N. M., ... & Collins, J. J. (2020). A deep learning approach to antibiotic discovery. *Cell*, 180(4), 688-702.
- Strickland, E. (2019). IBM Watson, heal thyself: How IBM overpromised and underdelivered on AI health care. *IEEE spectrum*, 56(4), 24-31.
- Szolovits, P. (2019). Artificial intelligence and medicine. In *Artificial intelligence in medicine* (pp. 1-19). Routledge.
- Tonekaboni, S., Joshi, S., McCradden, M. D., & Goldenberg, A. (2019). What clinicians want: contextualizing explainable machine learning for clinical end use. In *Machine learning for healthcare conference* (pp. 359-380). PMLR.
- Topol, E. J. (2015). *The patient will see you now: the future of medicine is in your hands* (Vol. 130). New York: Basic Books.
- Topol, E. J. (2019). High-performance medicine: the convergence of human and artificial intelligence. *Nature medicine*, 25(1), 44-56.
- Tran, B. X., Vu, G. T., Ha, G. H., Vuong, Q. H., Ho, M. T., Vuong, T. T., ... & Ho, R. C. (2019). Global evolution of research in artificial intelligence in health and medicine: a bibliometric study. *Journal of clinical medicine*, 8(3), 360.
- Tseng, P., Kaplan, R. S., Richman, B. D., Shah, M. A., & Schulman, K. A. (2018). Administrative costs associated with physician billing and insurance-related activities at an academic health care system. *Jama*, 319(7), 691-697.
- Tufail, A., Rudisill, C., Egan, C., Kapetanakis, V. V., Salas-Vega, S., Owen, C. G., ... & Rudnicka, A. R. (2017). Automated diabetic retinopathy image assessment software: diagnostic accuracy and cost-effectiveness compared with human graders. *Ophthalmology*, 124(3), 343-351.
- Vaishya, R., Javaid, M., Khan, I. H., & Haleem, A. (2020). Artificial Intelligence (AI) applications for COVID-19 pandemic. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, 14(4), 337-339.
- Vardazaryan, A., Mutter, D., Marescaux, J., & Padoy, N. (2018, September). Weakly-supervised learning for tool localization in laparoscopic videos. In *International Workshop on Large-scale Annotation of Biomedical data and Expert Label Synthesis* (pp. 169-179). Cham: Springer International Publishing.
- Wahl, B., Cossy-Gantner, A., Germann, S., & Schwalbe, N. R. (2018). Artificial intelligence (AI) and global health: how can AI contribute to health in resource-poor settings?. *BMJ global health*, 3(4), e000798.
- Walsh, C. G., Ribeiro, J. D., & Franklin, J. C. (2017). Predicting risk of suicide attempts over time through machine learning. *Clinical Psychological Science*, 5(3), 457-469.
- Weng, S. F., Reps, J., Kai, J., Garibaldi, J. M., & Qureshi, N. (2017). Can machine-learning improve cardiovascular risk prediction using routine clinical data?. *PloS one*, 12(4), e0174944.
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... & Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data*, 3(1), 1-9.
- World Health Organization. (2025). *Suicide worldwide in 2021: Global health estimates*. World Health Organization.
- Wu, E., Wu, K., Daneshjou, R., Ouyang, D., Ho, D. E., & Zou, J. (2021). How medical AI devices are evaluated: limitations and recommendations from an analysis of FDA approvals. *Nature Medicine*, 27(4), 582-584.
- Wynants, L., Van Calster, B., Collins, G. S., Riley, R. D., Heinze, G., Schuit, E., ... & van Smeden, M. (2020). Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. *bmj*, 369.

Zech, J. R., Badgeley, M. A., Liu, M., Costa, A. B., Titano, J. J., & Oermann, E. K. (2018). Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: a cross-sectional study. *PLoS medicine*, 15(11), e1002683.

Zhavoronkov, A., Ivanenkov, Y. A., Aliper, A., Veselov, M. S., Aladinskiy, V. A., Aladinskaya, A. V., ... & Aspuru-Guzik, A. (2019). Deep learning enables rapid identification of potent DDR1 kinase inhibitors. *Nature biotechnology*, 37(9), 1038-1040.