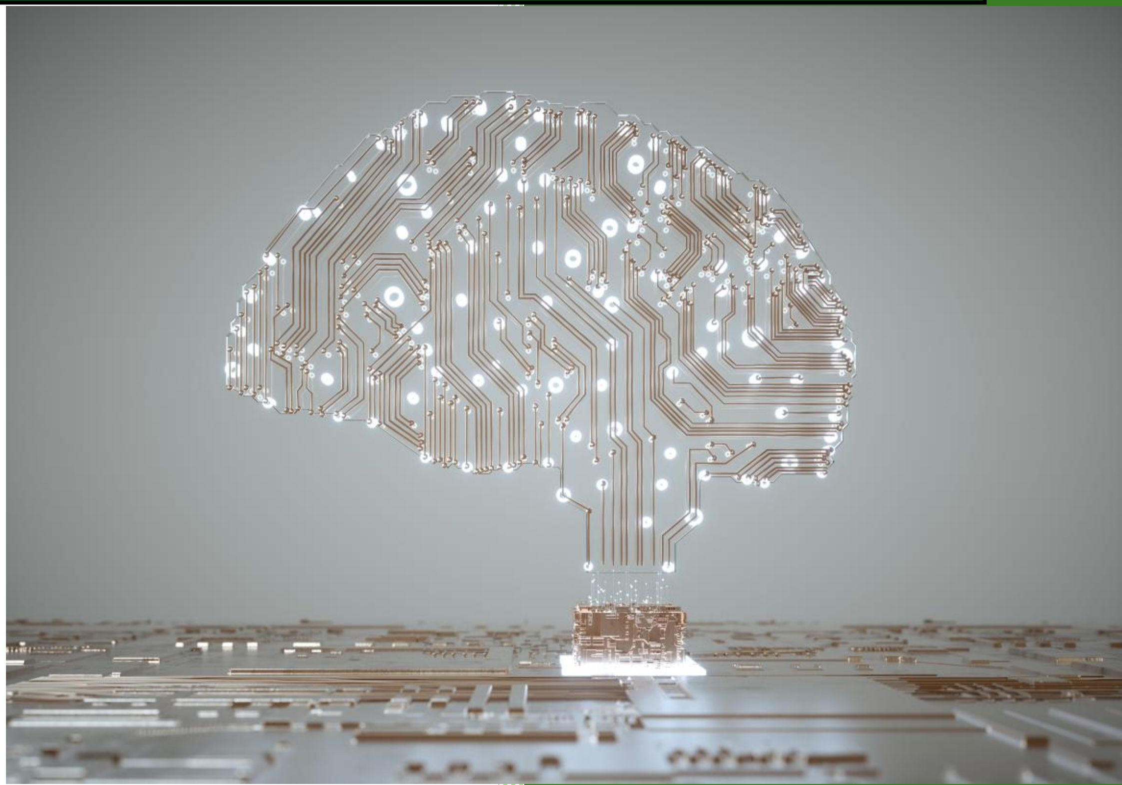
MBA-MPH Capstone



Amirali Yousefli

Integration of Large Language Models (LLMs) into Medical Practices: A Framework for Enhancing Diagnostic and Treatment Decisions in Medically Underserved Area

5/28/2025

Table of Contents

I. Introduction

Personal Motivation and Author's Healthcare AI Expertise Background

My interest in the intersection of healthcare and artificial intelligence stems from diverse experiences across pharmaceutical research, healthcare quality assurance, and health technology development. Through my current studies in the MBA/MPH program at Dartmouth, I've gained perspective on both the clinical and operational aspects of healthcare innovation—an interdisciplinary view that informs this research.

Working in pharmaceutical development and healthcare quality systems provided valuable insights into how healthcare solutions are created, validated, and delivered. These experiences revealed the significant challenges in maintaining quality while improving accessibility—a balance particularly difficult to achieve in resource-constrained settings.

More recently, working on AI healthcare applications has shown me the gap between technological capability and practical implementation. While developing systems that aim to support clinical decision-making, I've observed both the promising potential and substantial barriers to effective deployment, especially in underserved communities.

What motivates this research is the recognition of a critical disconnect: as healthcare AI advances rapidly, these innovations often fail to reach communities with the greatest need. Discussions with healthcare providers serving in resource-limited environments have highlighted how technological solutions frequently overlook the practical constraints they face daily. This observation has guided my focus toward developing implementation frameworks that address both technological and operational realities.

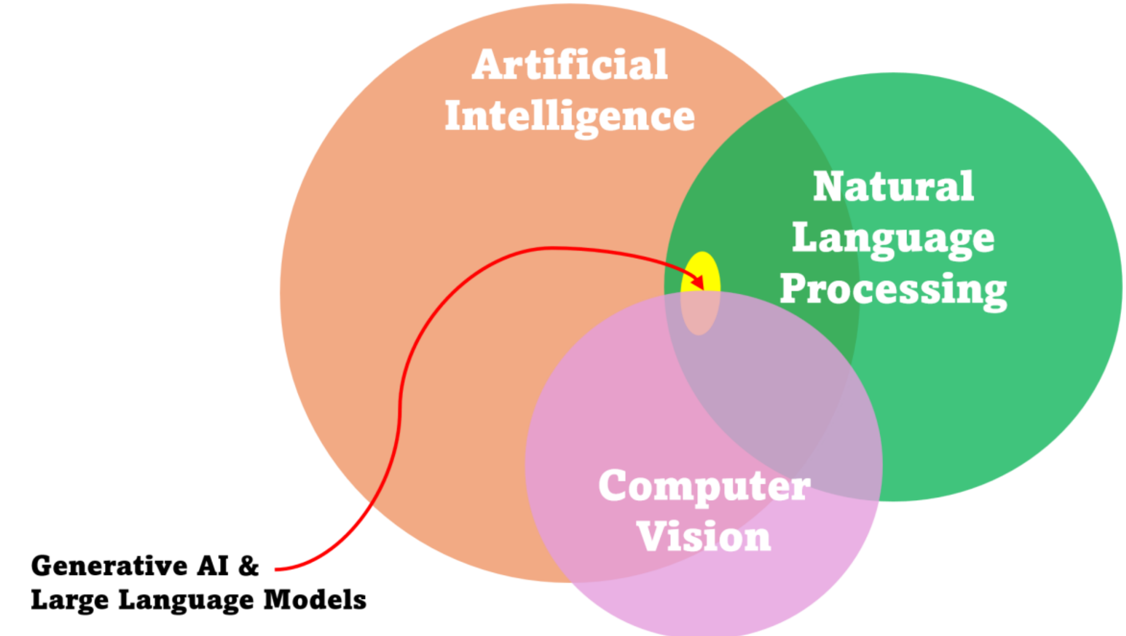
The MBA/MPH program has provided opportunities to engage with experts across healthcare, technology, and public health sectors, deepening my understanding of the multifaceted challenges in healthcare delivery. Each conversation and project has reinforced the importance of creating solutions that are not only technically sound but also operationally feasible and culturally appropriate for their intended settings.

This capstone represents an effort to contribute to the ongoing dialogue about responsible AI implementation in healthcare—specifically focusing on how these powerful tools might be thoughtfully adapted to address healthcare disparities rather than inadvertently widening them.

Definitions

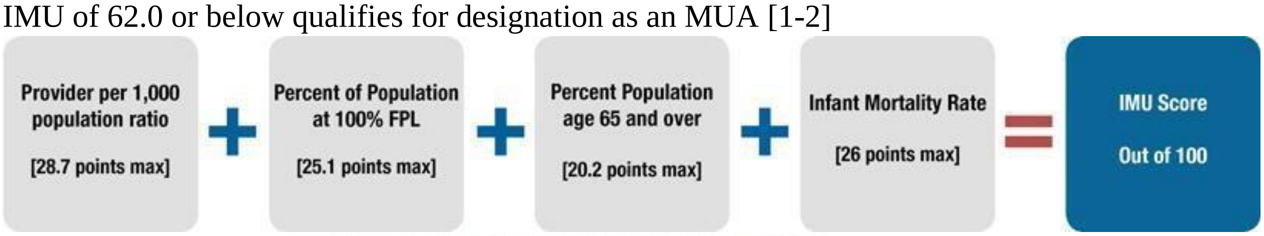
AI (Artificial Intelligence): The simulation of human intelligence in machines designed to learn, reason, and problem-solve, encompassing technologies from rule-based systems to advanced neural networks.
Generative AI: Artificial intelligence that is capable of generating new content (such as images or text) in response to a submitted prompt (such as a query) by learning from a large reference database of examples
NLP (Natural Language Processing): A field of computer science and linguistics that enables computers to understand, interpret, and generate human language in useful ways.

LLM (Large Language Model): A sophisticated neural network trained on vast text data that can understand context, generate human-like text, and perform various language tasks by identifying patterns and relationships between words and concepts.

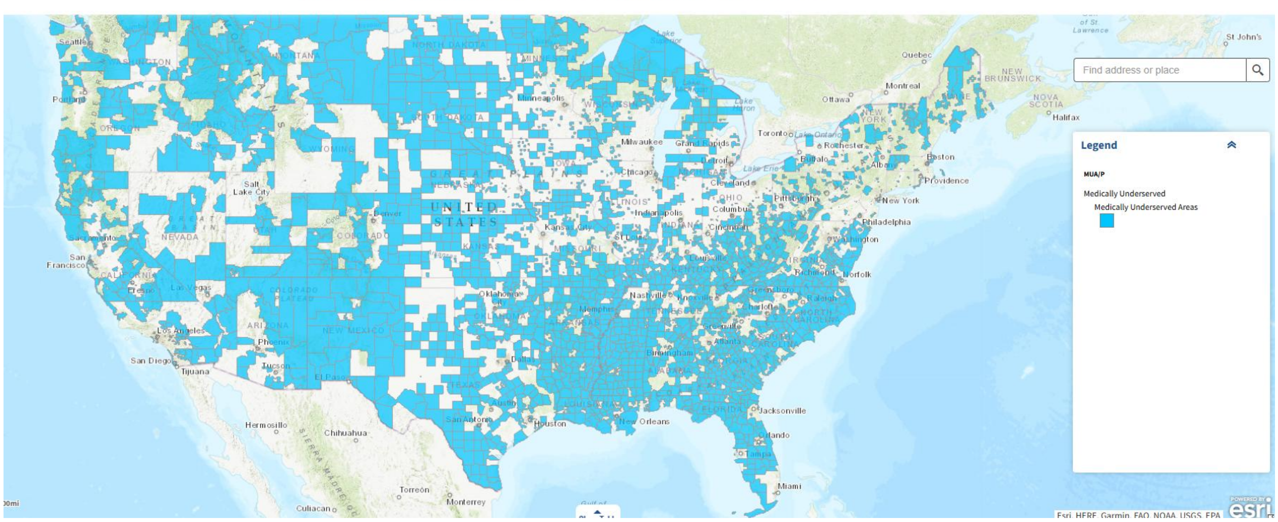


"Framework" : A structured methodology and set of interconnected principles that guides the systematic integration of Large Language Model diagnostic technologies into underserved healthcare settings.

Medically Underserved Area: EMUAs identify geographic areas and populations with a lack of access to primary care services. Under the established criteria, an area or population with an IMU of 62.0 or below qualifies for designation as an MUA [1-2]



Quick Maps - Medically Underserved Areas/Populations (MUA/P)



Research Methods

This research employs a mixed-methods approach combining primary and secondary research methodologies. Primary research includes 15 structured interviews with healthcare professionals, AI researchers, venture capitalists, and industry leaders across the healthcare AI ecosystem. Secondary research encompasses analysis of more than 9 healthcare AI companies and review of more than 10 peer-reviewed papers examining LLM applications in healthcare settings.

Problem Statement

Healthcare delivery in underserved regions faces critical challenges due to severe physician shortages, limited access to specialist expertise, and resource constraints. According to global health assessments, the most pronounced physician shortages occur in rural areas and developing countries, where physician-to-population ratios can be as low as 0.1 physicians per 1,000 people—twenty times lower than in developed nations [3]. This shortage translates into delayed diagnoses, limited treatment options, and ultimately poorer health outcomes for millions of people worldwide.

The problem is compounded by the uneven distribution of specialist expertise. Primary care providers in underserved settings often manage complex cases without timely access to specialist consultation, increasing the risk of diagnostic errors and suboptimal treatment plans. In many regions, the nearest specialist might be hundreds of miles away, requiring patients to undertake costly and time-consuming travel—a luxury many cannot afford. Infrastructure limitations further exacerbate these challenges, with inconsistent access to electricity, internet connectivity, and computing resources creating additional barriers to healthcare delivery.

Existing solutions like telemedicine initiatives, mobile health applications, and task-shifting approaches have shown promise but remain insufficient to close the gap. These approaches often fail to address the core issue: providing high-quality diagnostic support and clinical decision-making assistance to frontline healthcare workers who must operate independently in resource-constrained environments.

Gap between LLM potential and real-world healthcare implementation

Recent advances in Large Language Models present a compelling opportunity to address these challenges. LLMs have demonstrated remarkable capabilities in medical applications, from accurately interpreting complex clinical information to generating differential diagnoses that rival those of human physicians. The evolution of these models has been swift and impressive—progressing from basic natural language processing to sophisticated medical reasoning systems capable of integrating vast amounts of medical knowledge.

The transformative potential of AI extends beyond individual patient care to broader public health applications. AI technologies demonstrate substantial benefits across cardiovascular health, oncology, radiology, and critical care, with particular impact in low-resource settings [4]. Machine learning algorithms excel in automated image interpretation, predictive analytics, and surgical precision enhancement [4]. However, successful implementation requires addressing infrastructure constraints, data quality issues, and integration complexity—challenges that are magnified in underserved regions where healthcare worker shortages and limited technological resources create additional barriers.

As Dr. Asha Zimmerman's research has shown, even basic open-source models have demonstrated diagnostic performance comparable to human physicians when properly implemented. The potential for cost reduction is significant, with Dr. Zimmerman's estimation that a $3,000 emergency room visit could potentially be reduced to $300 with AI-assisted diagnostics. In regions where specialist consultation is scarce, these tools could provide critical decision support for frontline healthcare workers, potentially transforming care delivery in underserved settings.

However, a significant gap exists between the theoretical potential of this technology and its practical implementation in healthcare settings, particularly in underserved regions. As Dr. Saeed Hassanpour noted, despite widespread unofficial use of LLMs by healthcare professionals, formal implementation faces substantial challenges. These include:

1. **Technical Infrastructure Barriers**: Many underserved areas lack reliable internet and basic computing resources needed for AI deployment.
2. **Trust and Explainability Issues**: Healthcare providers and patients remain hesitant to rely on "black box" AI systems for critical medical decisions.
3. **Contextual Adaptation Challenges**: Models trained on data from advanced healthcare systems often perform poorly in different regions with unique disease patterns and resource constraints.
4. **Accountability Frameworks**: Unlike for human doctors, comprehensive frameworks for AI accountability in healthcare don't exist.
5. **Data Privacy Concerns**: Medical institutions are reluctant to share valuable patient data, complicating model development and implementation.
6. **Clinical Workflow Integration**: Challenges exist in seamlessly incorporating AI tools into existing healthcare practices.
7. **Business Viability Questions**: Sustainable monetization models for AI healthcare tools remain unproven, with several high-profile failures like Forward's AI-powered "CarePods."

The result is a concerning paradox: the communities that could benefit most from LLM-assisted healthcare delivery are precisely those least likely to access it without a deliberate, thoughtful implementation framework.

Evidence from systematic analyses reveals both the promise and perils of AI integration in healthcare. While AI enhances diagnostic accuracy and streamlines treatment planning across specialties like radiology and pathology, implementation faces significant hurdles including cultural resistance, regulatory challenges, and sustainability concerns [4]. The risk of overreliance on AI systems potentially reducing healthcare workers' critical thinking skills represents a particular concern in resource-limited settings where human expertise is already scarce [4]. These findings emphasize the critical need for thoughtful implementation strategies that augment rather than replace human clinical judgment.

Capstone Aim

This capstone project aims to develop an implementation framework for integrating LLM diagnostic support systems into healthcare settings in underserved regions. The framework seeks to bridge the gap between cutting-edge AI technology and practical healthcare delivery while addressing both clinical efficacy and business viability. Project objectives include:

1. Developing a structured approach for integrating LLM diagnostic support systems that accounts for varying levels of infrastructure readiness in underserved settings
2. Creating guidelines for ensuring clinical efficacy and safety in LLM implementation
3. Addressing technical infrastructure requirements through tiered deployment options
4. Establishing business models that balance sustainability with accessibility
5. Navigating complex and varying regulatory landscapes
6. Building trust mechanisms for both healthcare providers and patients

The urgency of developing such frameworks is underscored by rapid AI evolution and growing evidence of its healthcare applications. Recent studies demonstrate AI's capacity to improve patient outcomes across diverse medical specialties, from cardiovascular risk prediction to cancer detection and surgical precision [4]. However, the digital divide threatens to exacerbate existing healthcare disparities if AI implementation lacks deliberate focus on accessibility and equity [4]. This capstone addresses this critical gap by developing practical, evidence-based guidelines that ensure AI's transformative potential reaches underserved populations most effectively.

By creating this comprehensive framework, the project aims to provide a roadmap for healthcare organizations in underserved regions to effectively implement and leverage LLM technologies in their practice. The ultimate goal is to improve diagnostic accuracy, reduce healthcare disparities, and optimize limited medical resources while ensuring high standards of patient care and safety. Success would mean that millions of patients in underserved communities could benefit from improved healthcare delivery through responsible AI implementation—truly bringing the promise of this revolutionary technology to those who need it most.

II. Current State of Healthcare in Underserved Regions

Healthcare disparities: resource distribution and access challenges

Healthcare resources globally follow patterns of wealth and development, creating stark disparities between affluent urban centers and underserved regions. These disparities manifest in fundamental ways that profoundly affect patient experiences and outcomes. In many rural communities and developing regions, the journey to receive basic healthcare often involves hours or even days of travel, with patients navigating unreliable transportation systems and incurring significant costs that many simply cannot afford.

The urban-rural divide in healthcare access represents one of the most persistent disparities worldwide. While metropolitan areas concentrate healthcare facilities, specialists, and advanced technologies, rural residents face systematic barriers to care. Beyond mere distance, rural healthcare facilities typically offer fewer services, less specialized care, and operate with more constrained resources. This geographic maldistribution creates fundamentally different healthcare experiences based solely on location.

Socioeconomic factors further compound these disparities. In underserved regions, healthcare expenses often represent catastrophic financial burdens for families, forcing impossible choices between medical care and other essential needs like food or education. Many communities lack health insurance mechanisms or public safety nets, creating cycles where untreated conditions worsen until they reach crisis levels. As Dr. Mecchella observed, populations differ significantly across regions, with underserved areas often bearing heavier disease burdens while having fewer resources to address them.

The distribution of healthcare specialties follows similar patterns of inequality. While major medical centers might offer dozens of specialties and subspecialties, underserved regions often lack even basic specialty coverage in areas like psychiatry, orthopedics, or pediatrics. This absence creates care gaps that no amount of primary care alone can bridge, particularly for conditions requiring specialized diagnosis or treatment.

These disparities reflect deeply rooted structural inequities rather than random variation. Colonial legacies, economic policies, political priorities, and demographic changes have shaped health system development paths that systematically advantage some populations while marginalizing others. Understanding this historical context is crucial for developing solutions that address root causes rather than merely treating symptoms of healthcare inequality.

Key challenges: physician shortages, limited expertise, and infrastructure constraints

Physician shortages represent perhaps the most visible challenge in underserved healthcare settings. Beyond raw numbers, these shortages create cascading effects throughout health systems. Overburdened physicians must prioritize acute care, leaving little time for preventive services or chronic disease management. Patient visits become abbreviated, reducing opportunities for thorough examination and discussion. In many regions, primary care providers manage extraordinarily diverse caseloads without specialist support, increasing the risk of diagnostic delays and treatment errors.

Limited specialist expertise compounds these challenges. When specialists are scarce or absent, patients with complex conditions face impossible situations. As one rural physician described, "We become the cardiologist, the psychiatrist, the endocrinologist—not because we're qualified for those roles, but because there's simply no one else." Referral networks, taken for granted in well-resourced areas, often don't exist in underserved regions, forcing primary providers to manage conditions beyond their training or refer patients to facilities hundreds of miles away.

The integration of artificial intelligence in healthcare settings faces significant adoption barriers, particularly in underserved regions where technical infrastructure limitations compound existing challenges. Patient acceptance of AI-integrated healthcare applications depends heavily on performance expectancy, effort expectancy, and trust factors [5]. Privacy concerns and hedonic motivation (an emotional or pleasant experience that occurs due to technology use, and it may be either pleasure or happiness) also influence adoption decisions, while the COVID-19 pandemic has accelerated digital health technology acceptance among both providers and patients, potentially reducing traditional resistance to technological innovations in healthcare delivery [5].

Infrastructure constraints create fundamental barriers to healthcare delivery. Many facilities operate with unreliable electricity, inadequate water supplies, and insufficient medical equipment. Diagnostic capabilities considered basic elsewhere—laboratory testing, X-rays, ultrasound—may be unavailable or inconsistently accessible. Digital infrastructure limitations further isolate providers from clinical resources, continuing education, and consultation opportunities that could improve care quality. As Soroush Vosoughi noted, even newer, more efficient LLM models require basic computing resources and connectivity that cannot be taken for granted in many settings.

The healthcare workforce in underserved regions faces substantial education and training challenges. Medical education systems often concentrate in urban centers, with curricula designed for well-resourced environments that bear little resemblance to underserved settings. Continuing education opportunities are limited, making it difficult for providers to maintain and update their skills. Clinical isolation means fewer opportunities for case discussion, mentorship, and professional development that contribute to clinical expertise development.

Financial constraints affect every aspect of healthcare delivery in resource-limited settings. Health systems operate with inadequate budgets that force impossible tradeoffs between essential services. Supply chains for medications and equipment function unreliably, with frequent stockouts of basic supplies. Healthcare workers receive inadequate compensation, contributing to retention challenges and brain drain to better-resourced areas. At the patient level, out-of-pocket expenses create substantial barriers to care, even when services are nominally available.

Current solutions and their limitations (telemedicine, mobile health, task-shifting)

Various interventions have attempted to address healthcare disparities in underserved regions, yet significant limitations have prevented these approaches from closing the care gap.

Telemedicine initiatives offer theoretical promise for extending specialist expertise into remote areas. In practice, however, many telemedicine programs face substantial implementation challenges. Connectivity issues disrupt video consultations in precisely the areas most in need of remote support. Hardware limitations, from insufficient cameras for dermatological assessment to inadequate peripherals for thorough examination, restrict the scope of conditions that can be properly addressed. Cultural and language barriers may complicate remote provider-patient interactions, particularly for sensitive health concerns.

Digital health interventions must address the complex interplay between technological capabilities and user acceptance factors. Research demonstrates that effort expectancy significantly influences performance expectancy in healthcare technology adoption, while patient health engagement serves as a critical mediator between technological features and actual usage behavior [5]. Understanding these behavioral determinants becomes essential for designing sustainable digital health solutions that overcome traditional barriers to technology acceptance in underserved healthcare environments [5].

Mobile health applications have proliferated as smartphones become more widely available, yet their impact remains limited. Many applications assume literacy levels, technological familiarity, and health knowledge that doesn't align with underserved population realities. Data costs and device limitations restrict access to video-based content that might otherwise bridge literacy gaps. Most critically, mobile health solutions often address information barriers without resolving access barriers—as Sully F. Chen observed, when an AI diagnoses a condition but the patient has no access to specialists, the knowledge alone provides limited benefit.

Task-shifting approaches, where responsibilities traditionally performed by physicians are transferred to nurses, community health workers, or other cadres, have shown promise in expanding basic service coverage. However, these approaches face significant limitations in addressing complex cases requiring specialized knowledge and skills. Supervision and support systems for task-shifted workers often function inconsistently, raising quality concerns. Professional resistance and regulatory barriers frequently complicate implementation, with medical associations raising legitimate questions about scope of practice and patient safety.

Community health worker programs have demonstrated effectiveness for specific interventions but struggle with sustainability and integration challenges. Many programs operate as donor-funded projects with limited timeframes rather than as integral components of health systems. Training and supervision vary widely in quality and consistency. Recognition and compensation remain inadequate in many contexts, contributing to high turnover rates that undermine program stability and effectiveness.

These existing solutions, while valuable, have failed to close healthcare gaps for several fundamental reasons. Most importantly, they often address isolated aspects of complex, multifaceted problems without changing underlying structural inequities in resource distribution. Technology-focused interventions frequently fail to account for infrastructure limitations and social contexts in implementation settings. Many solutions originate from high-resource contexts and undergo insufficient adaptation for low-resource realities. Finally, sustainability challenges plague many interventions, with programs collapsing when external funding ends or initial champions move on.

Contemporary healthcare systems must navigate competing demands for both cost-effectiveness and individualized care delivery. Medical education increasingly emphasizes teaching healthcare efficiency principles through case-based discussions and cost-effectiveness analysis frameworks [6]. However, the challenge remains integrating population-level resource allocation decisions with patient-centered care approaches, particularly in underserved regions where scarcity necessitates difficult choices between economic sustainability and optimal individual patient outcomes.

This complex landscape of healthcare disparities, systemic challenges, and partial solutions creates both the opportunity and the imperative for innovative approaches like LLM integration. However, it also highlights the importance of designing implementations that specifically address the unique constraints and contexts of underserved healthcare settings rather than simply transferring technologies designed for resource-rich environments.

III. Evolution of LLMs in Healthcare

Technical development milestones: from early NLP to specialized medical models

The evolution of Large Language Models in healthcare represents one of the most significant technological progressions in medical informatics. This journey, marked by several distinct phases, has transformed how we conceptualize the role of artificial intelligence in healthcare delivery.

In 2017-2018, early applications of Natural Language Processing in healthcare emerged with the introduction of models like BERT (Bidirectional Encoder Representations from Transformers). These models primarily focused on basic text processing tasks such as medical document classification and information extraction from clinical notes. While groundbreaking, these early NLP applications lacked the contextual understanding necessary for complex medical reasoning.

The period from 2019 to 2020 saw significant advancement through transfer learning in biomedical NLP. As highlighted by Peng et al. (2019), researchers began evaluating models like BERT and ELMo across multiple biomedical benchmarking datasets. Their work demonstrated that models pre-trained on general domains could be effectively fine-tuned for specialized medical applications, establishing a critical foundation for future developments. This approach significantly improved performance on tasks like medical entity recognition and relationship extraction while requiring less domain-specific training data.

By 2020-2021, dedicated medical-specific language models emerged, trained specifically on medical literature, clinical notes, and healthcare datasets. These models incorporated domain-specific vocabularies and were fine-tuned to understand medical terminology, relationships, and concepts. The localization of language models to the medical domain significantly improved their performance on healthcare-specific tasks, marking a crucial transition from general-purpose AI to purpose-built medical tools.

The year 2022 witnessed the rise of foundation models for medical AI, as detailed by Moor et al. (2023) in Nature. These large-scale models, trained on vast and diverse datasets, demonstrated the ability to adapt to multiple healthcare tasks through prompt engineering and few-shot learning. Foundation models represented a paradigm shift, as they could perform reasonably well across various medical tasks without task-specific fine-tuning, demonstrating unprecedented flexibility.

The specialized medical LLM landscape expanded significantly in 2023 with the development of models like Google's Med-PaLM 2, which demonstrated remarkable capabilities in medical reasoning and diagnosis generation. Commercial entities like Jivi AI emerged with proprietary models reportedly outperforming general-purpose LLMs on medical benchmarks. These specialized models incorporated medical knowledge graphs and extensive training on clinical decision pathways, moving beyond simple pattern recognition toward medical reasoning.

The current frontier, spanning 2024-2025, has focused on multimodal integration and retrieval-augmented generation (RAG). Modern healthcare LLMs can now process and analyze multiple data types including text, images, and structured clinical data. As detailed by Ke et al. (2024), RAG approaches enhance model performance by incorporating real-time access to medical literature, clinical guidelines, and patient records. These innovations address previous limitations around factual accuracy and currency of medical knowledge, creating systems that can provide more reliable clinical support.

Current clinical applications: administrative support to diagnostic assistance

The practical applications of LLMs in healthcare have evolved from handling basic administrative tasks to supporting complex clinical decisions. This progression reveals both the expanding capabilities of these models and their increasing integration into healthcare workflows.

Early clinical applications in 2021-2022 focused primarily on documentation and administrative support. Systems like Microsoft's DAX Copilot automated clinical documentation, generating progress notes and patient summaries from doctor-patient conversations. According to survey data from 879 clinicians, these tools saved an average of 5 minutes per patient encounter, with 77% of clinicians reporting improved documentation quality. This initial focus on administrative burden reduction provided a low-risk entry point for AI in clinical settings while demonstrating tangible efficiency improvements.

The 2022-2023 period saw the emergence of more sophisticated clinical decision support systems. Epic's AI integration exemplifies this trend, with features like automated coding assistance and radiology follow-up tools that help clinicians interpret and act on clinical data. These systems began providing more substantive support for clinical decisions while maintaining a human-in-the-loop approach that emphasized augmenting rather than replacing clinician judgment.

By 2023, LLMs were being deployed for patient triage and symptom assessment. Applications similar to Vox Cura, developed by Dr. Zimmerman, demonstrated how AI could conduct structured patient interviews to gather clinically relevant information and provide preliminary assessments. These systems helped prioritize care based on urgency and connected patients with appropriate levels of service, potentially reducing unnecessary emergency department visits.

The 2023-2024 timeframe marked the transition to more advanced diagnostic assistance tools. Google's AMIE (Articulate Medical Intelligence Explorer) exemplifies this progression, demonstrating performance matching or exceeding primary care physicians in diagnostic accuracy during controlled studies with trained patient actors. These systems moved beyond simple triage to offer comprehensive diagnostic reasoning, considering complex symptom patterns and medical histories.

Current implementations in 2024-2025 have expanded to treatment planning and personalization. Tempus AI's platform, for instance, integrates patient genomic data with clinical information to help physicians make more informed treatment decisions, particularly in oncology. As noted by Kunal Nagpal from Tempus, their system connects with approximately 65% of US Academic Medical Centers and over 50% of US oncologists, demonstrating significant market penetration.

The integration with Electronic Health Record systems represents the latest frontier of clinical application. Epic's suite of AI tools now includes ambient charting used by 186 healthcare organizations and automated patient portal responses generating over 1 million message drafts monthly across 150 healthcare systems. This deep integration into existing clinical workflows marks a critical milestone in the normalization of AI as a standard component of healthcare delivery.

The umbrella review by Iqbal et al. identified six major themes in LLM healthcare applications: generating patient education materials, interpreting medical information, providing lifestyle recommendations, supporting medication decisions, offering perioperative care instructions, and optimizing doctor-patient interactions [7]. Notably, 82.4% of reviewed studies focused on general LLM usage in healthcare, while 17.6% explored specialized applications such as medical examinations and systematic review assistance [7]. This distribution reflects the current emphasis on broad applicability rather than narrow specialization.

Performance benchmarks compared to human clinicians

Assessing LLM performance against human clinicians provides crucial context for understanding their potential role in healthcare delivery. Several benchmark studies and expert observations offer insight into current capabilities and limitations.

Dr. Zimmerman's testing with real datasets demonstrated that even basic open-source models could achieve performance comparable to human physicians on diagnostic tasks. His team notably outperformed established scores on the MIMIC-IV benchmark, suggesting that properly implemented LLMs can match or exceed certain aspects of clinical reasoning. This finding is particularly significant for underserved settings where access to specialist expertise is limited.

A study cited by Dr. Evans provides additional evidence for LLM potential. The research found that properly prompted AI systems outperformed both physicians working without AI assistance and physicians working with AI. This suggests that optimal implementation may not simply involve augmenting clinician capabilities but potentially redesigning clinical workflows to leverage AI strengths while compensating for human limitations.

Commercial systems have reported impressive performance metrics, though these should be interpreted cautiously. Jivi AI's proprietary model has reportedly outperformed models like Google's Med-PaLM 2 and OpenAI's GPT-4 on the Open Medical LLM Leaderboard, suggesting rapid advancement in specialized medical AI capabilities. However, as Soroush Vosoughi noted, clinicians often prioritize explainability over pure performance, preferring slightly less accurate models if they can better understand their trustworthiness.

Important limitations remain in LLM performance. Sully F. Chen highlighted that despite advancements in language processing, LLMs perform poorly on medical images, operating at near "random guessing level" of accuracy. This is significant since many common medical queries involve visual assessment. Additionally, current models struggle to rapidly incorporate new medical information, potentially providing outdated recommendations when medical guidelines change.

Performance also varies significantly across different medical specialties and tasks. While LLMs demonstrate strong capabilities in conditions with well-defined presentations and established diagnostic criteria, they struggle with rare conditions, complex multisystem disorders, and patient populations underrepresented in training data. As Dr. Evans noted, medical practice varies substantially between regions, and AI trained on US medical data might perform poorly in different regions due to differences in disease patterns and cultural expressions of illness.

The contextual understanding of healthcare delivery represents another performance limitation. As James Feng observed, physicians don't simply provide standardized solutions but rather tailor recommendations based on patients' social, economic, and personal circumstances. Current LLMs struggle to replicate this pragmatic, context-aware approach to healthcare delivery, highlighting the continued importance of human judgment in clinical decision-making.

Privacy and ethical concerns represent critical barriers to LLM implementation in healthcare settings. Current publicly available models lack HIPAA compliance, preventing the sharing of protected health information [8]. Furthermore, LLMs may inadvertently expose patients to difficult medical information without appropriate emotional support mechanisms [8]. The conversational nature of these systems, while engaging, lacks the personalized care and contextual understanding that human clinicians provide during sensitive medical discussions [8]. These limitations necessitate careful consideration of implementation boundaries and oversight mechanisms.

Despite these limitations, the trajectory of performance improvement suggests increasingly viable clinical applications. Dr. Zimmerman predicted that the current "physician-in-the-loop" approach will eventually give way to more autonomous systems as performance metrics continue to improve. This evolution, however, will require addressing not only technical performance but also the complex challenges of trust, explainability, and accountability that currently limit clinical adoption.

IV. Implementation Challenges and Considerations

Technical infrastructure requirements: connectivity, deployment options, hardware needs

Implementing LLM solutions in underserved healthcare settings presents unique technical challenges that require thoughtful adaptation. Dr. Zimmerman's work with Vox Cura revealed that while minimal technical requirements include an internet connection and a computer, even these basics cannot be assumed in many underserved regions.

Connectivity solutions must address varying levels of internet access. In regions with reliable broadband, cloud-based deployment offers scalability and minimal local hardware requirements. However, many underserved areas face intermittent or low-bandwidth connectivity, necessitating alternative approaches. Edge computing solutions that run lightweight LLM versions locally with periodic updates can maintain functionality during connectivity gaps. In the most resource-constrained environments, SMS-based interfaces—as explored by Dr. Zimmerman's team—may provide the widest reach, particularly in regions where cell towers offer the primary connectivity.

Hardware requirements vary by deployment model. Cloud implementations require minimal local computing power but demand reliable internet. Edge deployments need more robust local hardware but offer greater resilience against connectivity issues. Recent developments in model efficiency, noted by Soroush Vosoughi, have enabled newer models like DeepSeek to run on basic laptops, potentially expanding deployment options. Looking forward, James Weinstein's suggestion of leveraging Starlink satellite connectivity could provide transformative infrastructure for remote regions, enabling more sophisticated LLM implementations without requiring extensive ground infrastructure.

Power supply reliability represents another critical consideration often overlooked in technology deployments. Inconsistent electricity access requires solutions like low-power hardware, battery backup systems, or solar charging capabilities to ensure system availability in clinical settings where power outages are common.

Clinical integration: workflow strategies, training requirements, patient education

Successfully integrating LLMs into clinical practice requires careful attention to existing workflows, training approaches, and patient education strategies. Dr. Mecchella emphasized that implementation must recognize what treatments are actually available in resource-limited settings—models designed for advanced healthcare environments often suggest diagnostic or treatment approaches that are impractical or impossible in underserved contexts.

Workflow integration strategies must balance efficiency improvements with disruption minimization. The success of Epic's AI tools and Microsoft's DAX Copilot demonstrates the importance of seamless integration with existing systems. In underserved settings, workflows are often already strained by resource limitations, making disruption particularly costly. Phased implementation approaches that begin with non-critical applications allow healthcare providers to build familiarity and trust with LLM systems before expanding to more sensitive clinical domains.

Training requirements extend beyond basic technical operation to include understanding LLM capabilities and limitations. Different provider types (physicians, nurses, community health workers) require tailored training approaches that address their specific roles and educational backgrounds. Frank Caicedo's work with Kenyan clinics highlights the importance of training recent healthcare graduates who may be more receptive to new technologies while offering competitive compensation to retain talent.

Patient education presents unique challenges in underserved settings where technological literacy may be limited. As Sully F. Chen noted, content must often be adapted to basic comprehension levels (e.g., "5th grade reading level"). Cultural acceptance of AI medical advice varies significantly across communities and age groups, necessitating context-specific education strategies that address local beliefs and concerns about technology in healthcare.

Trust and explainability: addressing physician and patient acceptance barriers

Trust represents perhaps the most significant barrier to LLM adoption in healthcare settings. Soroush Vosoughi identified that clinicians often prefer slightly less accurate models if they can better understand their trustworthiness and reasoning processes. This observation highlights the fundamental tension between model sophistication and explainability that implementation frameworks must address.

For physicians, trust barriers include concerns about model accuracy, accountability for errors, and professional autonomy. Implementation strategies should incorporate transparency about model limitations, clear accountability frameworks that delineate responsibilities between AI systems and human providers, and approaches that position LLMs as decision support tools rather than autonomous actors. Thomas Thesen emphasized that AI systems remain "black boxes" with non-deterministic responses, requiring continued human oversight for critical medical decisions.

Patient trust considerations differ significantly from provider concerns. As James Feng observed, patients might be particularly hesitant to rely on AI systems for important health decisions given their "black box" nature. Implementation frameworks must address patient concerns through appropriate disclosure about AI involvement in care, human verification of critical decisions, and clear communication about how AI tools support rather than replace human healthcare providers.

Explainability approaches vary in sophistication and effectiveness. Basic approaches include providing confidence scores alongside recommendations and citing medical literature to support conclusions. More advanced techniques involve generating natural language explanations of reasoning processes and visualizing key factors that influenced the model's assessment. However, as Vosoughi cautioned, current LLMs often generate convincing post-hoc justifications rather than revealing their true decision-making processes, creating a risk of false confidence in their reasoning.

Data challenges: quality issues, privacy concerns, bias mitigation

Data challenges in underserved healthcare settings span quality, privacy, and bias considerations. Dr. Hassanpour highlighted that medical institutions view patient data as valuable assets and show reluctance in sharing, creating asymmetric situations where institutions with rich data often lack computational capabilities while those with technical expertise struggle to access necessary medical data.

Data quality issues are particularly pronounced in resource-constrained environments where electronic record-keeping may be inconsistent or recently implemented. Missing data, non-standardized terminology, and handwritten records that resist digitization create significant challenges for training and implementing LLM systems. Implementation frameworks must include strategies for data preprocessing, quality assessment, and handling of incomplete information.

Privacy concerns take on different dimensions in underserved settings. While regulatory frameworks like HIPAA provide clear guidelines in the US, many regions lack equivalent protections. Dr. Mecchella noted that Dartmouth has implemented safeguards to prevent protected health information from being sent to public AI tools, but such institutional controls may be absent in resource-limited environments. Implementation approaches must therefore incorporate privacy-by-design principles and appropriate data governance structures regardless of local regulatory requirements.

Bias mitigation represents a critical ethical consideration. As Susan Dentzer highlighted through the Optum case study, algorithms/training data can inadvertently perpetuate existing biases by using proxies like past healthcare costs that reflect historical disparities. Dr. Evans emphasized that medical practice varies substantially between regions, with AI trained on US medical data potentially performing poorly in regions with different disease patterns and cultural expressions of illness. Implementation frameworks must include bias assessment, diverse training data where possible, and monitoring systems that detect performance disparities across different patient populations.

Regulatory considerations:

The regulatory landscape for healthcare AI varies dramatically across jurisdictions, creating complex compliance challenges for global implementations. Dr. Hassanpour explained that the FDA categorizes AI tools under Software as Medical Device (SaMD) classification, with approximately 200 AI tools having received FDA approval, primarily in medical imaging and cardiology. This established pathway provides guidance for US implementations but may not translate to other regions.

International regulatory variations create a patchwork of requirements. As Susan Dentzer noted, the European Union has enacted comprehensive AI legislation while the United States largely lacks federal regulations, with only a few state-level laws in place. Many developing regions have minimal AI-specific regulation, creating uncertainty about applicable standards and responsibilities. Implementation frameworks must incorporate regulatory mapping and compliance strategies that account for this diverse landscape.

Compliance strategies for underserved settings must balance regulatory rigor with practical implementation realities. Risk-based approaches that calibrate compliance requirements to the specific application can prevent regulatory burdens from blocking access to beneficial technologies. Pragmatic frameworks might differentiate between decision support tools that supplement provider judgment versus autonomous diagnostic systems that require more stringent oversight.

Risk management and accountability frameworks remain underdeveloped for AI in healthcare. Soroush Vosoughi identified this as a critical barrier, noting that unlike human doctors who can be held responsible through established mechanisms, there's no comprehensive framework for AI accountability. Implementation strategies must therefore define clear lines of responsibility, error reporting mechanisms, and processes for continuous improvement based on performance monitoring.

Sully F. Chen's observation that current models cannot rapidly incorporate new medical information highlights the regulatory challenge of maintaining compliance as medical knowledge evolves. Implementation frameworks should include update protocols that balance the need for current information with validation requirements and regulatory review processes. This challenge becomes particularly acute during situations like pandemics when protocols change rapidly and outdated guidance could harm patients.

V. Business and Operational Considerations

Cost structure and ROI analysis for underserved settings

Developing sustainable business models for LLM implementation in underserved healthcare settings requires unique approaches to cost structure and return on investment analysis. Unlike traditional healthcare technology deployments that assume robust infrastructure and payment systems, implementations in resource-constrained environments must reconsider fundamental assumptions about both costs and returns.

The cost structure for LLM implementations includes several key components that vary significantly based on deployment approach. Initial development costs can be substantial, particularly for specialized medical LLMs, but these are increasingly mitigated by the availability of open-source models and APIs. As Kunal Nagpal from Tempus explained, many organizations now focus on evaluating and fine-tuning existing LLMs rather than building proprietary models from scratch, significantly reducing upfront investment requirements.

Infrastructure costs represent a variable but critical component. Cloud-based deployments minimize initial hardware investments but require ongoing subscription fees and reliable internet connectivity. Edge computing approaches necessitate more substantial upfront hardware investments but can reduce long-term connectivity costs. For the most resource-limited settings, James Weinstein suggested that minimal infrastructure solutions leveraging basic mobile devices and satellite connectivity could support business models costing as little as "$0.10 a person" by eliminating middleware and unnecessary regulations.

ROI calculation in underserved settings must extend beyond traditional financial metrics to include public health impacts and healthcare system efficiencies. Dr. Zimmerman highlighted the potential for significant cost reduction, citing an example where a $3,000 emergency room visit could potentially be reduced to $300 with AI-assisted diagnostics. This system-level saving may not directly benefit the LLM provider but represents substantial value to healthcare systems and patients. Implementation frameworks should incorporate these broader economic benefits when making investment cases to potential funders and partners.

Timeframes for expected returns must reflect the realities of healthcare adoption cycles. Anish More's skepticism about business viability cited challenges in monetization and historical failures like Babylon Health. Forward's collapse after burning through nearly $400 million in investor funding demonstrates the risks of unsustainable business models that prioritize technological innovation over financial viability. Successful implementations will likely require patience capital that can support extended paths to profitability while prioritizing sustainable growth over rapid scaling.

Business Sustainability models: subscription approaches, freemium options, partnerships

Diverse sustainability models have emerged from early LLM healthcare implementations, offering valuable insights for underserved settings. These models must balance accessibility for resource-constrained environments with financial viability to ensure long-term sustainability.

Subscription-based approaches have shown promise in various contexts. Dr. Zimmerman's Vox Cura adopted a strategic pricing model with annual subscriptions at $30 or monthly at $9.99, targeting 1000 initial users to generate approximately $30,000 in annual revenue. This approach provides predictable revenue streams while maintaining affordability for individual providers. For institutional implementations, tiered subscription models that scale pricing based on facility size or patient volume can ensure equitable access while capturing appropriate value from larger organizations.

Freemium models offer particularly compelling opportunities for underserved settings. Dr. Zimmerman's business plan incorporated provisions for free access in developing regions while maintaining paid subscriptions in more affluent markets. This cross-subsidization approach allows healthcare providers in resource-constrained environments to access core functionality while premium features generate revenue from users with greater financial resources. Implementation frameworks should define clear boundaries between free and premium features that ensure essential diagnostic support remains accessible while creating value differentiation for paid tiers.

Public-private partnerships represent a critical sustainability mechanism for underserved regions. James Weinstein's work with the Gates Foundation demonstrates how partnerships between commercial entities and nonprofits can fund infrastructure development and ongoing operations while ensuring accessibility. Frank Caicedo's franchise model with Access Afya provides another partnership template, where a larger organization provides technology, training, and operational support while local clinics maintain day-to-day operations and community relationships. These partnership models distribute financial risk while leveraging the complementary strengths of different organizations.

Revenue diversification strategies can enhance sustainability by reducing dependence on direct user payments. Frank Caicedo noted that Kenyan clinic business models rely heavily on pharmaceutical sales, projected to generate 60-70% of revenue. Similarly, Tempus AI's business model uniquely combines clinical services with data monetization—each new patient sequencing represents both a profitable transaction and an addition to their valuable database. Implementation frameworks should identify analogous complementary revenue streams that align with local healthcare economics and regulatory considerations.

Income-tiered pricing structures can ensure both accessibility and sustainability across diverse economic contexts. These approaches set pricing based on ability to pay, whether defined by individual income, institutional budget, or national economic classification. Epic's multiple monetization approaches across its AI suite demonstrate how a single technical platform can support different pricing models for various market segments, from large academic medical centers to smaller community providers.

Market analysis: stakeholder needs and value propositions

Successful implementation requires nuanced understanding of diverse stakeholder needs and development of targeted value propositions that address specific pain points in underserved healthcare settings.

For healthcare providers, particularly those in resource-limited environments, key needs include diagnostic support for complex cases, efficient use of limited time, and guidance on appropriate resource allocation. The value proposition centers on extending capabilities without requiring extensive additional training or resources. As Dr. Mecchella noted, AI could help prevent "anchoring bias" by suggesting differential diagnoses that physicians might not consider—particularly valuable for primary care providers who regularly see common conditions but rarely encounter rare diseases. Implementation frameworks should emphasize these workflow improvements while recognizing the importance of maintaining provider autonomy and judgment.

Healthcare administrators and system leaders prioritize operational efficiency, quality improvement, and financial sustainability. The value proposition must demonstrate how LLM implementations can reduce costs through appropriate triage, minimize costly referrals, and optimize resource utilization. Epic's implementation of AI-driven coding assistants that ensure accurate billing and compliance offers a template for administrative value creation. In resource-constrained systems, demonstrating how LLM implementations can help identify patients needing early intervention or transfer to better-equipped facilities can resonate with administrators responsible for optimizing limited healthcare resources.

Patients in underserved settings seek improved access to quality care, reduced wait times, and appropriate treatment recommendations. The value proposition focuses on receiving timely, accurate health information and connecting with appropriate care levels without unnecessary travel or expense. As Kunal Nagpal described with Tempus's patient-facing application Olivia, successful implementations can empower patients throughout their care journeys by providing information and context about their medical records and treatment options. Patient education about AI's role in their care must be culturally appropriate and address legitimate concerns about technology-mediated healthcare.

Payers and funders, whether governmental agencies, insurance providers, or international aid organizations, focus on cost-effectiveness, scalability, and sustainable impact. The value proposition must demonstrate how LLM implementations can reduce overall healthcare costs while improving outcomes for target populations. Dr. Zimmerman's observation about the potential for significant cost reduction through appropriate triage and diagnosis could be particularly compelling for payers seeking to optimize limited healthcare budgets.

Regulatory bodies and healthcare authorities prioritize patient safety, data privacy, and healthcare quality standards. The value proposition must address how LLM implementations can improve care quality while maintaining compliance with relevant regulations and standards. As discussed in the regulatory considerations section, frameworks should emphasize transparent accountability structures and clear risk management approaches that align with both local and international standards where applicable.

Market timing considerations reveal both opportunities and challenges. Sully F. Chen outlined a gradual timeline for AI integration in healthcare—short term (productivity enhancer for clinicians), medium term (integration with existing systems), and long term (automating up to 95% of routine cases), with truly autonomous systems at least 10 years away due to regulatory hurdles. Implementation frameworks should align business models with this realistic adoption trajectory, recognizing that initial applications may focus on provider support rather than autonomous diagnosis, with corresponding implications for revenue generation and value creation.

VI. Case Studies and Key Lessons

Success Stories

Microsoft Dragon Copilot: Transforming Clinical Documentation Through Ambient AI

Microsoft's Dragon Copilot represents one of the most successful large-scale implementations of AI in healthcare, demonstrating how ambient listening technology can meaningfully address physician burnout while improving patient care. The solution combines Dragon Medical One's trusted dictation capabilities with DAX Copilot's ambient AI technology, which has assisted over 3 million ambient patient conversations across 600 healthcare organizations in the past month alone. This integration marks a significant evolution from basic voice recognition to sophisticated ambient intelligence that can capture, process, and transform multiparty clinical conversations into comprehensive specialty-specific documentation.

The scale of Microsoft's implementation provides compelling evidence of successful healthcare AI adoption. With 600,000+ clinicians across thousands of organizations now using the platform, DAX Copilot has generated over 1 million message drafts monthly across 150 healthcare systems. The technology's impact extends beyond simple efficiency gains—clinicians report saving 5 minutes per patient encounter, with 70% reporting reduced feelings of burnout and fatigue, and 62% stating they are less likely to leave their organization. Perhaps most significantly, 93% of patients report a better overall experience, suggesting that AI augmentation actually enhances rather than diminishes the human aspects of healthcare delivery.

Microsoft's success stems from its gradual, human-in-the-loop approach that positions AI as augmenting rather than replacing clinical judgment. The system requires physician approval for all generated content, maintaining professional autonomy while reducing administrative burden. Epic CEO Judy Faulkner noted that "ART's responses are often more empathetic than the very busy doctors," highlighting how AI can sometimes enhance the human elements of care by providing thoughtful, comprehensive responses that time-pressed physicians might not have the bandwidth to compose. The platform's integration with existing workflows through electronic health record systems like Epic has proved crucial, eliminating the need for healthcare providers to adapt to entirely new systems while gaining significant efficiency benefits.

The financial sustainability of Microsoft's model demonstrates the viability of ambient AI solutions. Rather than requiring massive upfront investments from healthcare organizations, the subscription-based approach allows institutions to realize immediate returns through improved physician productivity and retention. Northwestern Medicine reports that physicians using the solution see an average of 11.3 additional patients per month, with 24% less time spent on notes and a 17% decrease in "pajama time" working on administrative tasks late into the night. This combination of improved work-life balance and increased patient throughput creates sustainable value propositions for healthcare organizations while addressing critical workforce challenges.

Epic: Seamlessly Integrating AI Throughout the Healthcare Workflow

Epic's approach to healthcare AI highlights one of the most comprehensive integration of artificial intelligence capabilities into electronic health record systems, demonstrating how embedded AI can transform healthcare delivery without disrupting existing workflows. With AI seamlessly integrated into Epic's Electronic Health Record (EHR), the platform revolutionizes healthcare operations through a HIPAA-compliant pipeline that incorporates state-of-the-art language models like GPT-4. Epic's success lies not in replacing healthcare workers but in intelligently automating routine tasks while enhancing clinical decision-making capabilities.

The breadth of Epic's AI implementation showcases the potential for comprehensive healthcare transformation. Epic's MyChart in-basket augmented response technology (ART) is now in use at 150 healthcare systems and medical groups, generating 1 million drafts each month and saving clinicians about half a minute per message. Beyond message drafting, Epic has expanded AI applications to include automated medical coding assistance, radiology follow-up tools, and AI-driven charting that can queue up orders and diagnoses while populating flowsheets automatically for nurses. Epic contends that generative AI tools can help address some of healthcare's most urgent needs, from workforce burnout to staffing shortages, with studies showing that 40% to 60% of clinicians report experiencing burnout.

Epic's commitment to responsible AI implementation has established industry standards for healthcare AI governance and validation. The company released the industry's first open-source AI validation tool, called "seismometer," designed so any healthcare organization can use it to evaluate any AI model against local population data. This tool addresses critical concerns about AI bias and performance disparities across different patient populations, providing healthcare organizations with standardized evaluation criteria for any data source. The Health AI Partnership, including Duke Health, Mayo Clinic, and Kaiser Permanente, is using Epic's AI trust and assurance software suite to conduct studies and generate evidence around AI model validation.

The clinical impact of Epic's AI implementations demonstrates measurable improvements in both efficiency and care quality. UNC Health, one of Epic's pilot partners, has been able to push generative AI implementation "as fast as Epic will allow," with physicians reporting significant improvements in their ability to manage patient message volume. Epic's AI tools have proven particularly valuable in academic medical centers where the academic and research mission drives forward-thinking adoption of new technologies. The platform's ability to analyze medical records while looking for trends through natural language queries has transformed how clinical leaders explore data and identify patterns, moving from traditional reporting tools to conversational analytics.

Epic's business model demonstrates how established healthcare technology companies can successfully integrate AI capabilities while maintaining their core value propositions. Rather than creating entirely new products, Epic has embedded AI throughout its existing software suite, allowing healthcare organizations to realize AI benefits through familiar interfaces and workflows. With more than 1,000 vendor-created apps now live in Epic's ecosystem, including more than 200 new apps added in the past year, the platform has created a thriving marketplace for AI-enabled healthcare applications. This approach has enabled Epic to maintain its dominant position in healthcare technology while leading the industry's transition to AI-enhanced care delivery.

Tempus AI: Pioneering Precision Medicine Through Multimodal AI Integration

Tempus AI represents the successful application of artificial intelligence to precision medicine, demonstrating how comprehensive data integration and AI-driven insights can transform cancer care and expand into multiple therapeutic areas. Founded in 2015 by Eric Lefkofsky after his wife's breast cancer diagnosis, Tempus has built one of the world's largest libraries of clinical and molecular data, serving as an AI-enabled precision medicine platform. The company's success stems from its unique approach to combining genomic sequencing, clinical data analysis, and AI-powered decision support into an integrated platform that serves physicians, researchers, and patients.

Tempus's market penetration demonstrates the substantial demand for AI-driven precision medicine solutions. Approximately 65% of all Academic Medical Centers in the US are connected to Tempus, with over 50% of oncologists in the US connected through sequencing, clinical trial matching, and research-enabled partnerships. This extensive network creates powerful network effects—as more physicians use Tempus, the platform generates more patient data, which enhances its AI models and makes the service more valuable for additional physicians. The company has generated significant revenue growth, increasing from $321 million in 2022 to $532 million in 2023, with continued expansion into multiple therapeutic areas beyond oncology.

The clinical applications of Tempus's AI platform demonstrate how machine learning can enhance traditional diagnostic approaches through intelligent data integration. Tempus's "Intelligent Diagnostics" concept involves incorporating an individual patient's longitudinal phenotypic, morphologic, and molecular data, including outcome data from their EHR, to give laboratory test results clinical context. This approach has proven particularly valuable in oncology, where Tempus claims it can get a patient into a clinical trial in about two weeks through its TIME Trial Program, which uses clinical and molecular data to screen cancer patients against qualifying criteria and match them to applicable trials.

Tempus's business model demonstrates the viability of data monetization strategies in healthcare AI. The company operates through three complementary product lines—Genomics, Data, and Apps—that create network effects enhancing the overall value of its ecosystem. The Genomics line provides diagnostic testing that generates molecular data, the Data and Services line facilitates drug discovery through data insights and clinical trial matching, and the AI Applications line develops algorithmic diagnostics and clinical decision support tools. By operating both genomic testing and clinical data analysis under one platform, Tempus can layer patient information from its vast dataset on top of test results, yielding richer insights than competitors who focus on only one aspect.

Tempus's expansion beyond oncology illustrates how successful AI healthcare platforms can leverage their core capabilities across multiple therapeutic areas. The company has received

FDA clearance for its Tempus ECG-AF device that uses AI to help identify patients who may be at increased risk of atrial fibrillation, representing the first FDA clearance for an AF indication in cardiovascular machine learning-based notification software. Additionally, Tempus has developed pharmacogenomic testing for patients with depression and respiratory pathogen panels that can return results in hours versus days. This diversification demonstrates how AI platforms built around comprehensive data integration can create value across multiple medical specialties while maintaining their core value proposition of precision medicine.

Philips Healthcare: Advancing AI-Embedded Imaging Systems for Global Access

Philips Healthcare's integration of artificial intelligence into medical imaging systems showcases a successful model for embedding AI capabilities directly into medical devices while addressing global healthcare access challenges. Philips has developed the industry's first helium-free MRI technology through its BlueSeal systems, with over 1,500 BlueSeal MRI systems globally and 100+ AI applications integrated across their imaging portfolio. The company's approach combines hardware innovation with AI-powered software solutions to deliver advanced imaging capabilities in more accessible and sustainable packages.

The scale and impact of Philips's AI implementations demonstrate successful integration of artificial intelligence across multiple imaging modalities. Philips's SmartSpeed AI technology delivers up to 65% higher resolution or up to 3x faster scanning, compatible with 97% of clinical protocols. The company's AI applications span the entire imaging workflow, from automated exam planning through SmartExam to AI-enhanced image reconstruction and quantitative analysis. The Smart Reading technology provides cloud-based AI toolsets that integrate imaging and reading on MR scanners, offering fully automated generation of AI-based quantitative reports for neurological indications and prostate cancer.

Philips's sustainability and accessibility innovations address critical infrastructure challenges in underserved healthcare settings. The BlueSeal helium-free technology uses only 0.5% of today's standard helium amount, with more than 600 systems installed globally saving more than 1.5 million liters of helium since 2018. This innovation is particularly significant for underserved regions where helium logistics are challenging or costly. The BlueSeal MRI systems are up to 1,700kg lighter than conventional systems, enabling flexible installations in new locations including elevated floors or indoor sites with construction limitations.

The business model and clinical outcomes of Philips's AI-embedded imaging systems demonstrate sustainable approaches to healthcare technology innovation. Customers report significant efficiency gains, with one facility reducing exam times to less than 60 minutes per slot for multi-parametric whole-body exams, enabling them to scan 2 more patients per day. The combination of hardware efficiency improvements and AI-powered workflow optimization creates compelling value propositions for healthcare organizations seeking to increase patient throughput without compromising diagnostic quality. Philips has extended its BlueSeal technology to mobile MRI systems, expanding quality access to MRI exams for more patients in more places, demonstrating how AI-embedded medical devices can address healthcare access challenges in underserved regions.

Philips's approach to AI governance and validation provides a model for responsible AI implementation in medical devices. The company maintains transparency about AI functions through comprehensive documentation showing which solutions fall under EU AI definitions, demonstrating commitment to regulatory compliance and ethical AI principles. Healthcare professionals remain in control of all imaging procedures, with AI serving to enhance rather than replace clinical judgment. This balanced approach has enabled Philips to achieve broad adoption while maintaining safety standards and regulatory compliance across diverse healthcare environments.

Cautionary tale: Forward's implementation failures

Forward's ambitious attempt to revolutionize healthcare through AI-powered "CarePods" represents a significant cautionary tale for LLM implementation in healthcare settings. The company, which had raised $657 million and achieved unicorn status with a $1 billion valuation, positioned these self-serve medical kiosks as a solution for automating health screenings and diagnoses. Led by former Google executive Adrian Aoun, Forward envisioned transforming healthcare delivery by migrating traditional clinical functions to hardware and software solutions, with doctors positioned to handle only the most complex cases. However, just one year after launching CarePods with a $100 million Series E funding round, Forward announced its abrupt shutdown in November 2024, ceasing all operations and laying off nearly 200 employees.

The failure stemmed from fundamental misalignments between Forward's technology-centric approach and healthcare delivery realities. Former employees revealed significant technical challenges, including failed self-service blood draws and instances of patients getting stuck in pods. More critically, the company's core failure originated from its attempt to "productize" healthcare by removing human interaction from the equation—a fundamental misunderstanding of healthcare as an inelastic, service-based industry where personal connection remains essential. The $99 monthly subscription model targeting primarily healthy, wealthy urban millennials proved unsustainable when competing against traditional healthcare options, particularly given the high operational costs of maintaining both physical clinics and CarePod locations in expensive urban areas.

Forward's experience offers valuable implementation lessons for LLM integration in underserved settings. First, it demonstrates the risks of prioritizing technological innovation over human-centered healthcare delivery, suggesting that successful implementations should augment rather than attempt to replace human clinical judgment. Second, it highlights the importance of sustainable business models that align with healthcare market realities rather than venture capital growth expectations. Finally, it underscores the necessity for thorough validation of healthcare AI systems before scaling—Forward's rapid expansion before establishing clinical effectiveness and user acceptance contributed significantly to its ultimate failure. These lessons emphasize that LLM implementation frameworks must balance technological capabilities with healthcare delivery fundamentals, maintaining the human elements of care that technology alone cannot replicate.

Emerging hybrid approaches and regional adaptations

In contrast to Forward's technology-first approach, more successful implementations are embracing hybrid models that carefully balance AI capabilities with human expertise. Dr. Mecchella's perspective exemplifies this trend, viewing AI as a tool supporting physicians rather than replacing them, particularly valuable for preventing "anchoring bias" by suggesting differential diagnoses that physicians might not otherwise consider. Vox Cura's implementation in Kenya demonstrates this balanced approach, using sophisticated prompt engineering for structured patient interviews while maintaining healthcare provider oversight. Such hybrid models recognize both the strengths of LLMs in processing vast amounts of medical information and their limitations in contextual understanding and clinical judgment, creating implementations that enhance rather than attempt to replace human healthcare providers.

Regional adaptations addressing specific infrastructure limitations have emerged as essential for successful implementation in underserved settings. Frank Caicedo's work with Kenyan clinics highlights how implementations can be adapted to local resource constraints through partnerships with organizations like Access Afya, which provides comprehensive support including equipment, technology, and AI tools tailored to local conditions. Similarly, Dr. Zimmerman's exploration of SMS-based interfaces for regions with limited smartphone access demonstrates how technical approaches can be modified to match available infrastructure. James Weinstein's suggestion of leveraging Starlink satellite connectivity to create low-cost infrastructure solutions for approximately "$0.10 a person" represents an innovative approach to overcoming connectivity barriers in remote regions.

Cultural and contextual adaptations are proving equally crucial for effective implementation. Dr. Evans emphasized that medical practice varies substantially between regions, with AI trained on US medical data potentially performing poorly in regions with different disease patterns and cultural expressions of illness. This insight has led to implementations that incorporate local disease prevalence data, cultural expressions of symptoms, and available treatment options. Dr. Lisa Adams' caution against the "fallacy of the empty vessel" reminds implementers that patients enter healthcare encounters with their own belief systems, requiring LLM implementations to adapt to different cultural contexts and local health beliefs rather than imposing external frameworks.

Successful regional adaptations are increasingly incorporating community engagement as a core implementation component. Dr. Daniel Lucey's experience during disease outbreaks highlights the importance of finding trusted community representatives when introducing new healthcare technologies—"You have to find somebody that the community you're talking with trusts... somebody from their community." Implementations like Tempus's patient-facing application Olivia demonstrate how AI tools can be designed to empower patients within their specific healthcare contexts, providing information and support tailored to local needs and resources. These community-centered approaches recognize that technology adoption depends not only on technical performance but also on alignment with local values, practices, and healthcare delivery systems.

VII. Implementation Framework and Future Directions

Proposed framework components and adaptation guidelines

The proposed implementation framework for integrating LLM diagnostic support systems into underserved healthcare settings comprises five interconnected components designed to address the multifaceted challenges identified throughout this research. At its foundation lies the Infrastructure Assessment and Adaptation component, which evaluates available technical resources and establishes tiered deployment options ranging from cloud-based implementations for regions with reliable connectivity to edge computing solutions for intermittent internet access, and potentially SMS-based interfaces for the most resource-constrained environments. This assessment identifies the minimum viable infrastructure required for effective implementation while mapping potential upgrade paths as resources become available. As James Weinstein suggested, partnerships with satellite connectivity providers like Starlink could provide transformative infrastructure for remote implementations, potentially enabling sophisticated LLM deployments without extensive ground infrastructure.

The Clinical Integration and Validation component addresses the critical need for seamless incorporation into existing healthcare workflows while ensuring diagnostic accuracy and safety. Drawing from Dr. Mecchella's insights, this component emphasizes context-specific implementations that recognize what treatments are actually available in resource-limited settings rather than applying models designed exclusively for advanced healthcare environments. Implementations should begin with low-risk applications to build provider familiarity and trust before expanding to more sensitive clinical domains. Following Dr. Lisa Adams' recommendation, formal research protocols should evaluate patient satisfaction, diagnostic accuracy in local settings, clinical outcomes, and healthcare worker acceptance, with clear triggers for referring patients to in-person care when needed.

The Stakeholder Engagement and Training component establishes structured approaches for involving and educating all relevant parties, from healthcare providers to patients and community leaders. As Dr. Lucey emphasized, implementations should focus initially on younger clinicians who may be more receptive to technology while ensuring nurses—who provide most direct patient care—are involved early in the process. Training programs should address both technical operation and effective clinical integration, with content adapted to basic comprehension levels when necessary. Community integration through partnerships with local leaders, churches, and NGOs, as demonstrated in Frank Caicedo's Kenyan clinic model, helps establish trust and cultural acceptability while ensuring implementations address local priorities and concerns.

The Business Sustainability and Governance component outlines viable economic models adapted to different resource contexts while establishing clear accountability frameworks. Drawing from Dr. Zimmerman's Vox Cura model, this might include tiered subscription approaches with provisions for free access in the most resource-constrained regions subsidized by revenue from more affluent markets. Following Soroush Vosoughi's recommendation, implementation governance should establish transparent accountability mechanisms that clearly delineate responsibilities between AI developers, healthcare providers, and regulatory bodies. The Data Management and Privacy component completes the framework by establishing appropriate data practices for model training, validation, and ongoing improvement. This includes strategies for preprocessing inconsistently documented clinical information, implementing privacy-by-design principles regardless of local regulatory requirements, and incorporating bias assessment protocols that detect performance disparities across different patient populations. Together, these framework components provide a structured yet adaptable approach for implementing LLM diagnostic support systems that can be tailored to the specific needs and constraints of diverse underserved healthcare settings.

Technological developments: Starlink, offline capabilities, contextual awareness

Emerging connectivity solutions represent perhaps the most transformative technological development for LLM implementation in underserved regions. Traditional infrastructure limitations have severely constrained technology deployment in remote areas, but James Weinstein's work with satellite internet providers like Starlink demonstrates a potential leapfrog approach to connectivity challenges. These systems can deliver reliable internet to previously inaccessible locations without requiring extensive ground infrastructure—potentially enabling sophisticated LLM implementations in regions where traditional connectivity expansion would take decades. As Weinstein observed, this approach could support business models costing as little as "$0.10 a person" by eliminating middleware and traditional infrastructure requirements. Early implementations in collaboration with organizations like the Gates Foundation are already demonstrating the feasibility of this approach for healthcare delivery in remote regions.

Offline capabilities and edge computing solutions address the reality that even with expanding connectivity options, many healthcare settings will continue to face intermittent or unreliable internet access. Recent developments in model efficiency, as noted by Soroush Vosoughi, have enabled newer models like DeepSeek to run on basic laptops with increasingly modest hardware requirements. This trend toward more efficient models opens possibilities for robust edge deployments that can function effectively without constant cloud connectivity. As Sully F. Chen suggested, implementations could incorporate modular approaches where core diagnostic capabilities operate locally while more complex functions leverage cloud resources when available. Advances in model compression techniques, quantization, and hardware-optimized inference are steadily reducing the computational resources required for effective LLM operation, making truly offline medical decision support increasingly viable even in resource-constrained environments.

Contextual awareness represents a critical frontier for LLM implementation in diverse healthcare settings. As Dr. Evans emphasized, medical practice varies substantially between regions due to differences in disease prevalence, cultural expressions of symptoms, and available treatments. Next-generation medical LLMs are beginning to incorporate capabilities for regional adaptation through few-shot learning techniques that can rapidly adjust to local disease patterns and treatment protocols. Dr. Mecchella's observation that models designed for advanced healthcare settings often perform poorly in resource-limited environments highlights the importance of these adaptation capabilities. Future implementations will likely utilize localized medical knowledge bases that incorporate region-specific epidemiological data, available medications and diagnostics, and cultural factors affecting healthcare delivery. Some emerging systems already incorporate RAG approaches that reference locally maintained databases of treatment protocols and formularies to ensure recommendations reflect available resources.

These technological developments, when combined, could fundamentally transform healthcare delivery in underserved regions. A system leveraging Starlink connectivity for initial deployment and regular updates, edge computing for daily operation without consistent connectivity, and contextual awareness features for region-specific recommendations could provide diagnostic support in previously inaccessible settings. As models become more efficient and hardware requirements decrease, implementation costs will likely continue to fall, further expanding accessibility. However, as Dr. Lisa Adams cautioned, these technical advances must be accompanied by careful attention to cultural factors and local knowledge systems to ensure acceptance and effectiveness. The most promising technological trajectory appears to be one that combines increasing technical capability with greater adaptability to local contexts—providing powerful decision support while respecting the unique conditions and constraints of each implementation setting.

Impact potential: access improvements, cost savings, quality enhancement

LLM implementation in underserved healthcare settings offers transformative potential for improving healthcare access across multiple dimensions. By extending diagnostic support to regions with severe physician shortages, these systems can help address one of the most fundamental barriers to healthcare delivery. Frontline providers equipped with LLM support could manage a wider range of conditions with greater confidence, reducing the need for patients to travel prohibitive distances for specialist consultation. As Dr. Lucey observed during his work in epidemic settings, even basic diagnostic support for conditions like dehydration assessment could significantly improve care when "no X-rays, no blood tests, no intravenous fluids" are available. The asynchronous nature of many LLM implementations further expands access by enabling healthcare workers to consult these systems during patient encounters without requiring simultaneous specialist availability. For patients in the most resource-constrained environments, the difference could be fundamental—receiving timely, informed care locally versus going without care entirely due to distance, cost, or provider limitations.

The economic impact of properly implemented LLM systems extends beyond direct technology costs to healthcare system efficiency and resource optimization. Dr. Zimmerman highlighted the potential for significant cost reduction, citing examples where a "$3,000 emergency room visit could potentially be reduced to $300 with AI-assisted diagnostics." This cost difference reflects not just reduced specialist time but more appropriate resource utilization through improved triage and diagnostic accuracy. By helping frontline providers identify which patients truly require referral to higher-level facilities, LLM systems could significantly reduce unnecessary transportation costs and specialist consultations while ensuring those who need advanced care receive it promptly. For resource-constrained healthcare systems and populations without insurance coverage, these efficiency improvements translate directly to expanded care capacity and reduced financial barriers to access. Even modest improvements in resource allocation could yield substantial benefits in settings where healthcare budgets are severely limited and every resource must be carefully optimized.

Quality enhancement through LLM implementation manifests in several dimensions, from improved diagnostic accuracy to more consistent adherence to clinical best practices. Dr. Mecchella emphasized that AI could help prevent "anchoring bias" by suggesting differential diagnoses that physicians might not consider—particularly valuable for primary care providers who regularly see common conditions but rarely encounter rare diseases. The ability to access current medical knowledge at the point of care addresses a significant challenge in underserved settings where providers often practice in isolation with limited opportunities for continuing education or specialist consultation. As Thomas Thesen noted, while AI systems remain "black boxes" requiring human oversight for critical decisions, their ability to process vast amounts of medical information offers valuable complementary capabilities to human providers. Additionally, LLM systems could enhance care consistency through improved adherence to clinical guidelines and protocols, addressing the variability in care quality that often disproportionately affects underserved populations.

The cumulative impact of these access improvements, cost savings, and quality enhancements could significantly reduce healthcare disparities between well-resourced and underserved regions. While a single technology cannot address all structural inequities in healthcare delivery, thoughtfully implemented LLM systems represent a promising approach to extending limited healthcare resources and expertise. As Dr. Zimmerman's research demonstrates, even basic open-source models can achieve diagnostic performance comparable to human physicians when properly implemented. This suggests that the technical foundations for meaningful impact already exist—the critical challenge lies in developing implementation frameworks that address the multifaceted challenges of deployment in resource-constrained environments. By approaching implementation with attention to both technological capabilities and contextual realities, LLM systems could help bridge the gap between the theoretical potential of medical AI and practical improvements in healthcare delivery for historically underserved populations.

VIII. Conclusion and Recommendations

Key findings summary and critical success factors

This research reveals that implementing Large Language Models in underserved healthcare settings represents both a significant opportunity and a multifaceted challenge requiring balanced consideration of technical, clinical, cultural, and economic factors. The analysis of expert interviews and case studies demonstrates that LLMs have rapidly evolved from basic administrative support tools to sophisticated systems capable of diagnostic reasoning and clinical decision support. As Dr. Zimmerman's research indicates, even basic open-source models have demonstrated diagnostic performance comparable to human physicians on certain benchmarks. However, the path from technical capability to meaningful healthcare improvement in underserved settings is neither straightforward nor guaranteed. The stark healthcare disparities documented in underserved regions—characterized by severe physician shortages, limited specialist access, and resource constraints—create both the imperative for innovative solutions and the challenging implementation context that any framework must address.

The investigation of implementation challenges across technical, clinical, trust, data, and regulatory dimensions reveals several interconnected barriers that must be simultaneously addressed for successful integration. Technical infrastructure limitations require tiered deployment approaches ranging from cloud implementations to offline capabilities depending on local connectivity. Clinical integration demands careful workflow adaptation and provider training tailored to different professional roles and backgrounds. Trust and explainability challenges necessitate transparency about system limitations and clear accountability frameworks that delineate responsibilities between AI systems and human providers. Data challenges span quality issues in underresourced settings to privacy concerns and bias mitigation requirements. The regulatory landscape adds further complexity through its inconsistency across jurisdictions, requiring context-specific compliance strategies. The cautionary tale of Forward's implementation failure underscores the consequences of neglecting these multidimensional challenges in favor of technology-first approaches that misunderstand healthcare's fundamentally human and service-oriented nature.

Critical success factors for LLM implementation in underserved settings emerge clearly from this analysis. First, successful implementations must embrace hybrid approaches that position AI as augmenting rather than replacing human healthcare providers, recognizing both the strengths of LLMs in processing medical information and their limitations in contextual understanding. Second, implementations require robust community engagement strategies that involve local stakeholders from initial planning through deployment and evaluation, addressing Dr. Lucey's observation that technology acceptance depends on finding trusted community representatives. Third, business models must balance sustainability with accessibility through approaches like tiered pricing or cross-subsidization between markets, avoiding the unsustainable venture capital expectations that contributed to Forward's failure. Fourth, implementations must incorporate ongoing evaluation and adaptation mechanisms that assess not only technical performance but also practical clinical utility, patient outcomes, and healthcare worker acceptance—following Dr. Lisa Adams' recommendation for formal research protocols that evaluate multiple dimensions of impact. Finally, contextual adaptation represents perhaps the most crucial success factor, requiring systems to incorporate local disease patterns, available treatments, cultural expressions of illness, and healthcare delivery realities rather than imposing models designed exclusively for advanced healthcare systems. Together, these success factors form the foundation for responsible and effective LLM implementation that can meaningfully contribute to addressing healthcare disparities rather than inadvertently reinforcing existing inequities.

Targeted recommendations for different stakeholders

Healthcare Providers: Frontline clinicians should approach LLM integration with a balanced perspective that recognizes both the technology's potential and its current limitations. Begin with low-risk applications like documentation assistance and information retrieval before progressing to more complex clinical support functions, establishing comfort and familiarity with the systems through incremental adoption. Actively participate in system evaluation by providing structured feedback on diagnostic suggestions, workflow integration challenges, and patient reactions to AI-assisted care. Following Dr. Mecchella's observation that many cases don't fit standard patterns, maintain appropriate clinical skepticism and override system recommendations when they conflict with clinical judgment or local contextual factors. Finally, contribute to developing local knowledge bases by documenting region-specific disease presentations, treatment outcomes, and resource constraints to improve the contextual relevance of LLM recommendations.

Healthcare Administrators and System Leaders: System leaders should establish clear governance structures for AI implementation that operate at the highest organizational levels rather than delegating oversight exclusively to IT departments. As emphasized in the Healthcare

Standards Institute's governance framework, AI oversight should report directly to senior leadership or board level given its strategic importance and potential impact on clinical outcomes. Develop phased implementation plans that begin with non-critical applications to build organizational capacity and stakeholder acceptance before expanding to more sensitive clinical domains. Allocate resources for comprehensive training programs that address not just technical operation but also effective clinical integration and appropriate override protocols. Implement robust monitoring systems that track both technical performance metrics and practical impacts on workflow efficiency, provider satisfaction, and patient outcomes, creating feedback loops for continuous improvement.

Technology Developers and Vendors: Developers must prioritize creating flexible systems that can adapt to varying infrastructure capabilities in underserved settings, following Sully F. Chen's recommendation to build modular solutions rather than all-encompassing systems. Incorporate region-specific customization capabilities that allow adjustment for local disease prevalence, available medications, and cultural factors affecting healthcare delivery, addressing Dr. Evans' observation that medical practice varies substantially between regions. Invest in developing robust explainability features that help providers understand the basis for system recommendations, addressing the trust barrier identified by Soroush Vosoughi as fundamental to clinical adoption. Consider developing tiered pricing structures or cross-subsidization models that ensure accessibility in resource-constrained environments while maintaining business sustainability through revenue from better-resourced markets, following Dr. Zimmerman's approach with Vox Cura.

Policymakers and Regulators: Regulatory bodies should develop flexible, risk-based frameworks specifically adapted for underserved healthcare contexts that balance safety requirements with the need for innovation and accessibility. Following Susan Dentzer's observation about regulatory disparities between regions, work toward greater international harmonization of AI healthcare regulations to reduce compliance burdens for global implementations. Establish clear guidelines for human oversight requirements that vary based on clinical application risk, creating appropriate safeguards while avoiding unnecessarily restrictive approaches that might prevent beneficial implementations in resource-constrained settings. Develop specific regulatory pathways for AI systems designed to address healthcare disparities, potentially including expedited reviews or modified requirements for implementations in officially designated underserved areas, similar to existing incentive programs for other healthcare services in these regions.

Funding Organizations: Investors, foundations, and aid organizations should adopt "patient capital" approaches for healthcare AI implementations that allow for extended development timelines and realistic paths to sustainability rather than expecting rapid scaling and returns. Following Graham Brooks' insight about leveraging traditional healthcare channels rather than direct-to-consumer approaches, prioritize funding models that work through established healthcare delivery organizations rather than attempting to create entirely new care delivery systems. Require robust impact evaluation plans that measure not only technical performance but also practical clinical outcomes and accessibility improvements for underserved populations. Support implementation research alongside technology development, recognizing that effective deployment strategies are as critical to success as the underlying technical capabilities, particularly in resource-constrained environments.

Researchers and Academics: The research community should prioritize developing standardized evaluation frameworks for LLM implementations that include metrics for technical performance, clinical utility, cost-effectiveness, and health equity impacts. Expand research beyond major academic medical centers to include diverse implementation settings, addressing the limited geographic diversity of current evidence noted in the limitations section. Conduct longitudinal studies that track both short-term implementation outcomes and longer-term impacts on healthcare access, quality, and outcomes in underserved settings. Develop interdisciplinary research collaborations that combine clinical, technical, anthropological, and economic perspectives to address the multifaceted challenges of LLM implementation in diverse healthcare contexts, ensuring that technical solutions are informed by deep understanding of local healthcare delivery realities.

Community Leaders and Representatives: Community representatives have a critical role in building trust and ensuring cultural appropriateness of LLM implementations. Advocate for community involvement from the earliest planning stages rather than after key decisions have been made, ensuring implementations address local priorities and concerns. Following Dr. Lucey's insight that effective communication requires trust through community representatives, help identify trusted local figures who can introduce and explain new healthcare technologies to community members. Participate in developing culturally appropriate explanations of AI's role in healthcare that address local beliefs and concerns about technology while establishing realistic expectations about capabilities and limitations. Monitor implementation impacts on different community segments, particularly vulnerable groups, to ensure equitable access and benefit distribution and to identify any unintended consequences requiring adjustment.

Vision for equitable healthcare access through responsible AI implementation

The integration of Large Language Models into healthcare delivery represents not merely a technological advancement but a potential transformation in how medical expertise is distributed globally. Where healthcare resources have historically concentrated in affluent regions, properly implemented LLM systems could extend diagnostic support and clinical guidance to settings that have long operated with severe resource constraints. This vision is not about replacing human healthcare providers but about amplifying their capabilities—enabling a nurse in rural Kenya to access diagnostic reasoning comparable to specialists at major medical centers, or helping a community health worker in a remote village identify patients requiring urgent intervention. When designed with equity as a foundational principle rather than an afterthought, these systems could help dismantle longstanding barriers to healthcare access that have persisted despite decades of traditional interventions.

Responsible implementation requires balancing technological capabilities with deep respect for local contexts and human relationships in healthcare delivery. The framework developed through this research envisions implementations that adapt to local infrastructure realities, incorporate regional disease patterns and available treatments, engage communities meaningfully in planning and deployment, and establish clear accountability mechanisms that maintain human judgment in clinical decision-making. This balanced approach recognizes both the transformative potential of medical AI and the complex social, cultural, and economic contexts in which healthcare is delivered. It represents a middle path between uncritical techno-optimism that ignores implementation challenges and defensive resistance that misses genuine opportunities for improvement.

The economic dimension of healthcare access cannot be separated from technological capability. LLM implementations hold promise for not just extending clinical expertise but also optimizing limited healthcare resources through improved triage, reduced unnecessary referrals, and more consistent adherence to evidence-based practices. As Dr. Zimmerman observed, the potential cost differential between an emergency room visit and AI-assisted diagnosis could fundamentally alter healthcare economics in resource-constrained environments. These efficiency improvements translate directly to expanded access—more patients receiving appropriate care within existing resource constraints—while sustainable business models ensure that implementations can persist beyond initial funding periods to create lasting healthcare infrastructure.

Looking forward, the successful implementation of LLMs in underserved healthcare settings could contribute to a more equitable global health system where access to quality care depends less on geographic location or economic status. This vision does not naively suggest that technology alone can solve deeply rooted structural inequities in healthcare delivery. However, it recognizes that thoughtfully implemented LLM systems could form an important component of broader health system strengthening efforts—particularly in regions where specialist shortages will persist for decades despite traditional training approaches. By developing implementation frameworks that specifically address the unique challenges of underserved settings, we can help ensure that this powerful technology serves as a force for reducing rather than reinforcing healthcare disparities, bringing us closer to the goal of quality healthcare as a reality for all people rather than a privilege for the fortunate few.

IX. References

[1]. HRSA, “What is Shortage Designation? | Bureau of Health Workforce,” bhw.hrsa.gov, Jun. 2023. https://bhw.hrsa.gov/workforce-shortage-areas/shortage-designation

[2]. “Quick Maps,” data.hrsa.gov. https://data.hrsa.gov/maps/quick-maps?config=mapconfig/MUA.json

[3]. A. Honda et al., “For more than money: willingness of health professionals to stay in remote Senegal,” Human Resources for Health, vol. 17, no. 1, Apr. 2019, doi: https://doi.org/10.1186/s12960-019-0363-7.

[4]. Ravi Rai Dangi, A. Sharma, and Vipin Vageriya, “Transforming Healthcare in Low-Resource Settings With Artificial Intelligence: Recent Developments and Outcomes,” Public Health Nursing, vol. 42, no. 2, Dec. 2024, doi: https://doi.org/10.1111/phn.13500.

[5]. D. E. Kalisz, “Drivers of artificial intelligence integrated healthcare applications. Patients’ perspective,” Technological Forecasting and Social Change, vol. 216, pp. 124144–124144, Apr. 2025, doi: https://doi.org/10.1016/j.techfore.2025.124144.

[6]. L. A. Bock, S. Vaassen, W. N. K. A. van Mook, and C. Y. G. Noben, “Understanding healthcare efficiency—an AI-supported narrative review of diverse terminologies used,” BMC Medical Education, vol. 25, no. 1, Mar. 2025, doi: https://doi.org/10.1186/s12909-025-06983-5.

[7]. U. Iqbal, A. Tanweer, A. R. Rahmanti, D. Greenfield, L. T.-J. Lee, and Y.-C. J. Li, “Impact of large language model (ChatGPT) in healthcare: an umbrella review and evidence synthesis,” Journal of Biomedical Science, vol. 32, no. 1, May 2025, doi: https://doi.org/10.1186/s12929-025-01131-z.

[8]. J. A. Omiye, H. Gui, S. J. Rezaei, J. Zou, and R. Daneshjou, “Large Language Models in Medicine: The Potentials and Pitfalls,” Annals of Internal Medicine, vol. 177, no. 2, Jan. 2024, doi: https://doi.org/10.7326/m23-2772.

X. Appendices

Appendix 1: Key Opinion Leaders Interviewed

Name	**Title, Organization**	**Timeline**
Thomas Thesen, PhD	Associate Professor, Geisel School of Medicine at Dartmouth	Nov 2024
Dr. Arthur Evans	Professor of Medicine, Weill Cornell Medical College	Nov 2024
Anish More	Senior Director, Healthcare Services, RA Ventures	Nov 2024
James Feng	MD/MPH Candidate, Geisel School of Medicine at Dartmouth	Nov 2024
Saeed Hassanpour, PhD	Founding Director, Dartmouth Center for Precision Health and AI	Jan 2025
Soroush Vosoughi	Assistant Professor of Computer Science; Technical Associate Director, Center for Precision Health & AI	Jan 2025
Frank Caicedo	Hybrid MPH Candidate/Aspiring Healthcare Entrepreneur (Kenya Health Clinic)	Jan 2025
Asha M Zimmerman, MD	Assistant Professor of Surgery	Feb 2025
John Mecchella	Associate Professor of Medicine, Geisel School of Medicine at Dartmouth	Feb 2025
Jim Weinstien	Senior Vice President, Microsoft Healthcare	March 2025
Kunal Nagpal	Machine Learning Lead, Tempus AI	March 2025
Sully F. Chen	AI Research Scientist, OpenAI	March 2025
Graham Brooks	Partner, .406 Ventures	April 2025
Daniel Lucey, MD, MPH	Clinical Professor of Medicine, Geisel School of Medicine at Dartmouth	April 2025
Lisa Adams, MD	Associate Dean for Global Health, Geisel School of Medicine at Dartmouth	April 2025

Appendix 2: Use of Artificial Intelligence Tools

This document was produced with assistance from artificial intelligence tools: Claude 4 & 3.7 Sonnet Extended Thinking, GPTo3, o4-mini-high, Gemini 2.5 Pro, elicit.com, and Perplexity. The LLMs were used to help generate ideas, provide research suggestions, and assist with editing and proofreading. However, all content was carefully reviewed, fact-checked, and revised by the human author. Any errors or omissions remain the sole responsibility of the author. The use of AI was intended to enhance the research and writing process, not to replace human judgment or academic rigor.