



HARVARD Kennedy School

MOSSAVAR-RAHMANI CENTER
for Business and Government

Harnessing Artificial Intelligence to Advance Value-Based Contracting in Health Care

Roshini Moodley Naidoo
Harvard Kennedy School

Mohammed Y Safirulla Karamathulla
Harvard Kennedy School

Amine Mernissi
Harvard Kennedy School

May 2025

M-RCBG Associate Working Paper Series | No. 258

The views expressed in the M-RCBG Associate Working Paper Series are those of the author(s) and do not necessarily reflect those of the Mossavar-Rahmani Center for Business & Government or of Harvard University. The papers in this series have not undergone formal review and approval; they are presented to elicit feedback and to encourage debate on important public policy challenges. This paper is copyrighted by the author(s). It cannot be reproduced or reused without permission. Pursuant to M-RCBG's Open Access Policy, this paper is available to the public at hks.harvard.edu/centers/mrcbg free of charge. Papers may be downloaded for personal use only.

Roshini Moodley Naidoo
Mohammed Y Safirulla Karamathulla
Amine Mernissi

Harnessing Artificial Intelligence to Advance Value-Based Contracting in Health Care

The US Health Care System is extraordinarily advanced, with a lattice of providers, payers, policy-makers, and intermediaries. The traditional fee for service (FFS) model, by which most care is funded in the nation, is recognized as a significant contributor to health care costs, incentivising volume over value. Value-Based Contracting (VBC) has been implemented as an alternate funding model to FFS, where value is a function of risk-sharing between payers and providers targeting quality and cost improvements. Arguably an over-simplification, and perhaps even an over-reach of what is practically possible within medium-term horizons, the defining VBC principles still provide a strategic and operational compass to guide strategy. In the US, the Affordable Care Act (ACA 2010) fostered a prioritization of Value-Based Care, where VBC programs introduced as part of the ACA, linked provider reimbursement to performance on quality and cost. While aspirational, VBC models continue to face challenges in conceptualisation and implementation, contributing to slower than expected adoption.

This Working Paper is structured in two parts. The first part deals with the evolution of VBC in the US and provides an analysis of a Medicare program – the Merit Based Incentive Payment System (MIPS) which offers physicians financial incentives based on their performance on selected metrics. Though MIPS is not a VBC program in its purest form, it provides an opportunity to understand implementation challenges when shifting from FFS towards VBC that impact adoption of the latter. The second part of the paper explores the opportunities presented by emerging AI technologies by examining four Use Cases. We assess the potential of AI technologies to improve the dual goals of VBC – increasing quality and reducing cost. Additionally, our paper examines the potential of AI solutions to increase the shift to VBC strategies by mitigating implementation challenges.

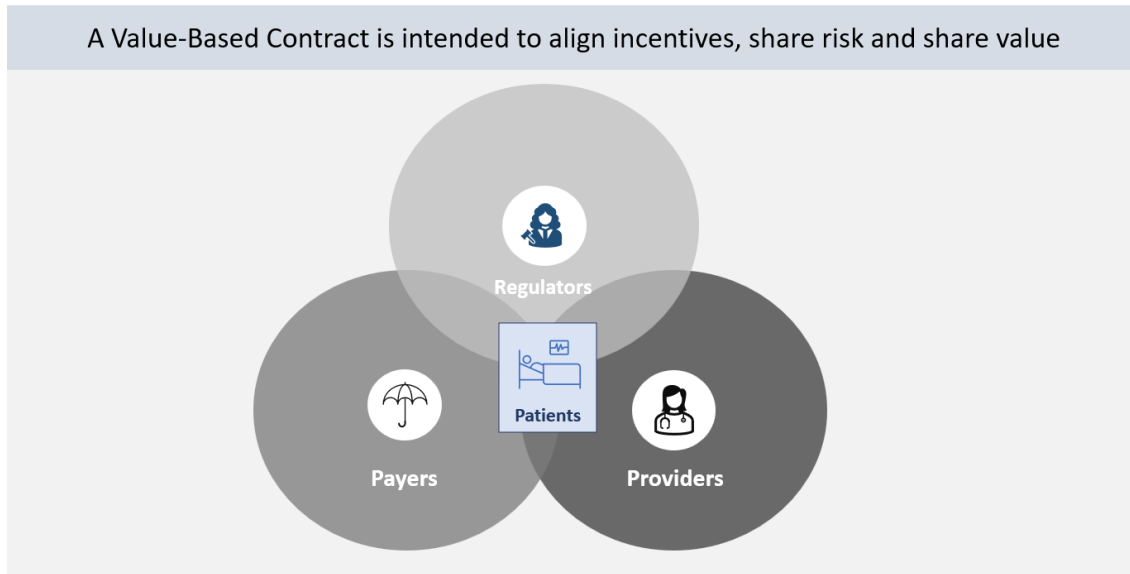
PART I. Value-Based Contracting

A. Background

The concept of Value-Based Contracting (VBC) dates back to the mid-2000's. As early as 2003, CMS launched the Premier Hospital Quality Incentive Demonstration, where Medicare incentives were developed based on a pay for performance model¹. VBC has advanced since then with the overarching objective to promote better stakeholder alignment (providers, payers, regulators and patients) for improved health care costs and outcomes of care.

¹ Centers for Medicare & Medicaid Services (CMS). Premier Hospital Quality Incentive Demonstration. Accessed May 28, 2025. <https://www.cms.gov/medicare/quality/initiatives/hospital-quality-initiative/premier-quality-incentive-demonstration>

Figure 1. Alignment of Stakeholder Goals in Value-Based Contracting



To more clearly illustrate the dynamics of the VBC payment mechanism, we examine at conceptual level, a sentinel VBC program launched in 2017 by CMS: The Merit-based Incentive Payment System. The MIPS program targets a wide range of disciplines (including physicians, osteopathic practitioners, chiropractors, nurse practitioners, and physical therapists) and provides bonus payments and penalties based on providers’ performance². A provider’s performance is defined according to several criteria including **quality of care outcomes** and the **cost of delivering these outcomes**. Clinicians report 6 quality measures from a list of 200 options. Each measure is scored from 1 to 10 points based on a comparison with national benchmarks, and payments are adjusted by the score obtained. On the cost front, a clinician’s performance is measured based on Medicare claims data which is compared with national benchmarks, and the performance score is then reflected in the payment adjustments. Weights have varied over time to quality and costs having costs have the same weight when CMS determines the final payment bonus/ penalty received by clinicians³. In 2023, the Health Care Payment Learning and Action Network (HCPLAN) reported 61.6% of payments in the US (92.7% of all payments) in Alternative Payment Models, however only 28.5% of these payments demonstrated down-side risk, the latter reflecting the intent of the shared risk principles of VBC⁴.

Health care authorities initially expressed ambitious adoption targets for VBC. In 2015, Health and Human Services Secretary Sylvia Burwell shared her objectives: “Our goal is to have 85% of all Medicare fee-for-service payments tied to quality or value by 2016, and 90% by 2018”⁵. Despite

²Centers for Medicare & Medicaid Services (CMS). How MIPS Eligibility Is Determined. Published September 2024. Accessed April 11, 2025. <https://qpp.cms.gov/mips/how-eligibility-is-determined>

³Centers for Medicare & Medicaid Services (CMS). Traditional MIPS Scoring Guide for the 2024 Performance Year. Published June 2024. Accessed April 11, 2025. <https://qpp-cm-prod-content.s3.amazonaws.com/uploads/2875/Traditional-MIPS-Scoring-Guide-for-the-2024-Performance-Year.pdf>

⁴Health Care Payment Learning and Action Network (HCPLAN). APM Measurement, Progress of Alternative Payment Models. 2024 Methodology and Results Report

⁵Sylvia M. Burwell. Setting Value-based payment goals - HHS efforts to improve US Health Care. The New England Journal of Medicine. 2019; DOI: 10.1056/NEJMp1500445

a steady increase in adoption, these targets were not met. Fee-for-service payment schemes with no link to quality or cost performance still represented 40% of insurance payments in 2018⁶. Several factors, mostly linked with VBC implementation, explain the mismatch between the expressed ambitions and actual adoption.

B. VBC implementation

The implementation of VBC raises several challenges that affect adoption. Some have to do with data limitations - the efficient deployment of VBC requires the development of robust technological assets and capabilities. Others are related to administrative costs - for example, complying with a VBC program's requirements. Equally, performance monitoring is time consuming and onerous for many health care providers.

The MIPS program illustrates several VBC implementation challenges that have near universal relevance to VBC strategies more widely. The challenges described below, are particularly salient for the impact on VBC adoption:

- Performance data accessibility: CMS issues an annual MIPS Feedback Report to physicians, providing detailed insights of performance across key metrics. In the absence of real-time performance feedback, the annual MIPS report loses much of its utility in helping medical practitioners improve performance. More often the Report serves as a retrospective audit analysis rather than a proactive improvement tool for performance enhancement⁷.
- Administrative burden: The reporting requirements for MIPS are time-consuming and resource intensive. A qualitative study determined that a physician spends an average of approximately 53 hours on MIPS-related activities per annum. The cost to participate in the program is therefore significant and is estimated at \$13,000 per year, while the bonus payment received as part of MIPS is often lower⁸.
- EHR interoperability: Deployment of information technology solutions is costly, and maintaining them as per the required technical standards requires significant maintenance costs. Few hospitals have IT systems at superior levels of sophistication; in rural areas and in many independent non-profit hospitals, limited phases of patient interaction exist in digital form⁹. Even in places where digital solutions have been implemented, IT systems operate at various stages of iterative technology upgrades. This poses a significant challenge to interoperability between various IT systems with a resultant impact on quality and cost of care.
- Information technology for quality improvement: Related to EHR interoperability above, providers differ in the capabilities and competencies required to undertake quality improvement efforts to meet MIPS performance requirements. Therefore, even where providers are enabled with information technology capabilities, these capabilities have to be leveraged in ways that improve specific quality metrics. Additionally, patients may visit a physician as little as once or twice a year for a chronic condition (often associated with payer visit rules). Very little information (if anything) is available to the physician pertaining to

⁶Rachel M. Werner, Ezekiel J. Emanuel et al. The Future of Value-Based Payment: A Road Map to 2030. Penn LDI Leonard Davis Institute of Health Economics. 2021

⁷American Medical Association (AMA). Letter to Chiquita Brooks-LaSure, CMS Administrator, Regarding MIPS Cost Performance Category. October 27, 2023. Accessed April 11, 2025. <https://searchf.ama-assn.org/letter/documentDownload?uri=%2Fstructured%2Fbinary%2Fletter%2FLETTERS%2Ffcls.zip%2F2023-10-27-Letter-to-Brooks-LaSure-re-MIPS-Cost-Performance-Category-v2.pdf>

⁸Hockenberry JM. Time and Financial Costs for Physician Practices Participating in the Medicare Merit-based Incentive Payment System: A Qualitative Study. JAMA Health Forum. Published February 10, 2021. Accessed April 11, 2025. <https://jamanetwork.com/journals/jama-health-forum/fullarticle/2779947>

⁹Turbow S, Hollberg JR, Ali MK. Electronic health record interoperability: How did we get here and how do we move forward? JAMA Netw Open. Published 2021. Accessed April 29, 2025

disease control between visits, with the added complexity of patients consulting different physician during these periods. In sum, at practice level, health care professionals have limited resources to implement substantive quality improvement strategies.

Recognizing the potential of AI as a VBC enabler, CMS in 2024 issued a Request for Information on Artificial Intelligence Technologies for Improving Health Care Outcomes and Service Delivery, and asked healthcare organizations to share uses of AI with potential to improve care quality¹⁰. In Part 2 of our paper, we highlight the potential for AI technologies to influence the dual goals of VBC i.e. improving quality and reducing cost, while simultaneously mitigating the implementation challenges that hinder adoption. We anticipate that as AI technologies advance, a more pressing question is likely to be whether AI solutions can lead to superior results in quality and costs than traditional non-AI approaches to VBC.

PART II. Leveraging AI to enable Value-Based Contracting by improving Cost and Quality

A. Proposed framework for assessing the potential of AI solutions to enhance Value-Based Contracting

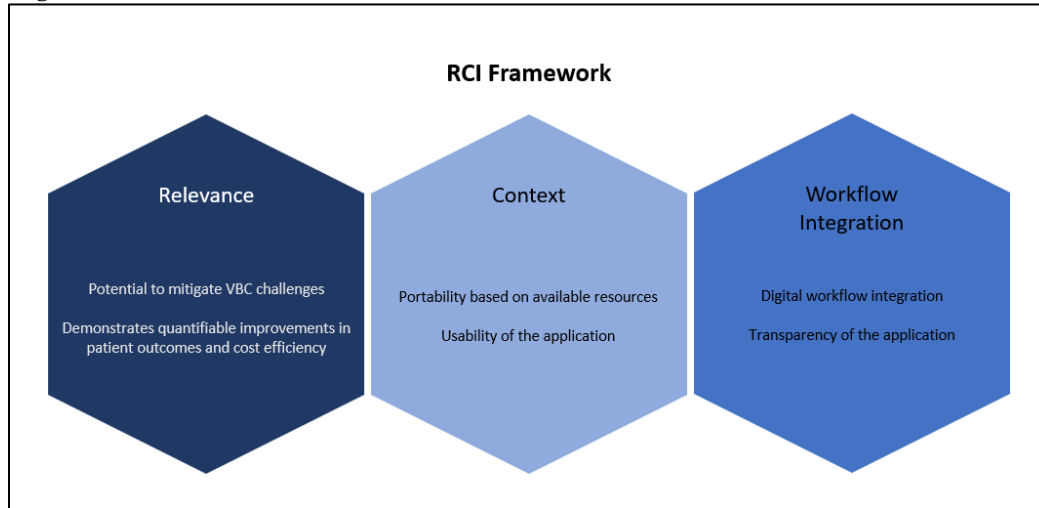
Given the nascency of most AI solutions being deployed in health care, there is a paucity of evidence evaluating the effectiveness and portability of AI solutions across settings, operating context, and populations. We propose a framework for analysing the opportunities presented by AI applications to enhance Value-Based Contracting (VB), involving three dimensions:

- Relevance
- Context
- Integration

We apply the framework to systemically evaluate AI solutions for each of the challenges associated with VBC adoption challenge presented by MIPS (as the VBC example). The forthcoming analysis describes the framework and applies it to assess the potential for AI solutions to mitigate VBC challenges. The framework assumes that the AI solution has been scientifically and technically validated. Additionally, we use the Framework to examine the potential of the AI solution to improve quality and reduce costs, thereby testing the potential of the innovation to meet the goals of VBC. It is worth noting, that beyond the cost and quality of care, AI solutions may incur unique investment (e.g. training, security and scalability) and benefits (increased workforce productivity).

¹⁰Centers for Medicare & Medicaid Services (CMS). Artificial Intelligence Demo Days. Published November 5, 2024. <https://www.cms.gov/digital-service/artificial-intelligence-demo-days>

Figure 2. Dimensions of the RCI Framework



Relevance (R)

It is fair to expect that for AI solutions to enable VBC, the technology should substantively reduce the burden posed by the current value-based care ecosystem by mitigating the factors identified as barriers to adoption. The Relevance dimension examines the AI solution from two sub-perspectives.

Potential to mitigate key VBC challenges

While it would be overly ambitious to expect AI solutions to solve all VBC implementation challenges, to enhance VBC, any AI solution should effectively address at least one of the major barriers to VBC implementation described in Part 1 (using MIPS as an example):

- Performance data accessibility
- Administrative burden
- Interoperability
- Information technology for quality improvement

Demonstrates quantifiable improvement in quality of care and cost efficiency

At its core, VBC is defined by the goal to achieve the best outcomes in care combined with eliminating wasteful costs. It follows, that AI solutions, to function as VBC enablers, must similarly demonstrate the potential for improved patient outcomes and cost efficiencies. Arguably, as for traditional VBC strategies, AI technologies should not be required to achieve symmetry between outcome and cost gains, preferencing outcomes over costs as guided by VBC goals. Robust evidence of quality and cost benefits should be provided to allow comparison with the incumbent non-AI strategy. Additionally, evaluation of the AI solution should include specific metrics that align with at least one of the key performance categories traditionally associated with VBC – Clinical Quality Measures (CQM), Cost Efficiency Measures, Patient Experience Measures, and Service Provider Performance Improvement Measures.

Context (C)

A key barrier to widescale VBC adoption is that the operating circumstances and resources required for implementation may differ vastly between providers and settings where care is delivered. The Context dimension examines the AI solution from two sub-perspectives below.

Portability based on available resources

It is essential that the AI application is adaptable across varied settings for example, rural and urban, small provider practices and expansive ACO's. A rigid solution which matches narrowly with only certain healthcare operating ecosystems, while potentially offering attractive benefits for those systems, may widen divisions between settings in which care is delivered.

Usability of the application

The AI solution must be designed in such a way that the health care workforce can easily and seamlessly be trained on its use. A consideration that requires ongoing interrogation, is the accessibility of AI solutions to patients, both physically and conceptually. It remains to be better understood how patients with differing literacy levels, and varying preferences for technology, will impact AI adoption rates.

Integration (I)

Clinical workflows, defined as the activities and processes pertaining to history taking, physical examination, diagnosis, treatment, referral, and follow-up, are the bedrock of encounters between patients and their physicians. Work that occurs outside an established workflow is generally burdensome to providers and results in unwarranted variation from standardized, evidence-based clinical protocols. This dimension examines the ease of integration of the technical functions of an AI solution within clinical workflows. The Integration dimension examines the AI solution from two sub-perspectives.

Digital workflow integration

EHR's are fast becoming the standard by which clinical pathways are implemented in practice. The AI solution must be able to integrate with the provider's EHR, extending to the property of interoperability where this exists. Therefore, it is prudent that AI solutions support standardized and widely recognized data formats such as Health Level Seven International (HL7) and Fast Healthcare Interoperability Resource (FHIR). Similarly, applications that are not vendor neutral increase platform dependency and limit the potential for future scalability.

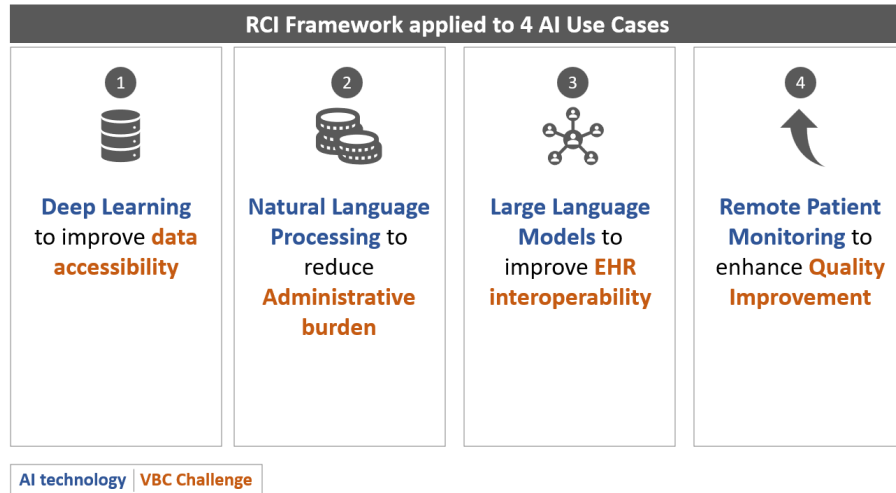
Transparency of the application

As described elsewhere, transparency is an essential requirement for AI health care applications. The underlying logic on which the AI application is based must be translatable to stakeholders, including patients should they be interested. Given the vulnerability to bias, documentation pertaining to the model design should be freely available for independent validation. Transparency can be further enhanced through the availability of training datasets, demographic composition, model architecture, performance metrics (for example, accuracy, sensitivity and specificity), and the selection and weighting of features. This is particularly important in health care, where trust is crucial to influencing health care behaviors of patients that in turn are correlated with care outcomes.

B. AI Enablement of Value-Based Contracting - Use Cases

Our focus below is the application of the "Relevance, Context, Integration" (RCI) framework to assess the potential for AI solutions to enable VBC, through four distinct AI Use Cases. Each Use Case examines the potential to mitigate a key VBC challenge (as described in Part 1 using MIPS as the example). We examine the potential for AI innovations to improve quality and reduce costs - the central tenet of VBC, and anticipate that this focus will grow as AI technologies mature.

Figure 3. Application of the RCI Framework to AI Use Cases



1. Deep Learning to improve Performance data accessibility

Delayed reporting is a long-standing and near-universal challenge facing VBC models. Lags in how data are shared between providers and payers make progress towards targets difficult to monitor. More importantly, improvements that providers may implement to achieve VBC targets may equally be difficult to measure. Providers are left feeling disempowered, not knowing what to fix, and when fixes are actually implemented, what the result of these are. Deep Learning Models can help in addressing this challenge. IBM defines Deep Learning as "a subset of machine learning that uses multilayered neural networks, called deep neural networks, to simulate the complex decision-making power of the human brain"¹¹. Deep Learning models offer a solution to performance data lags by providing insights at the point at which care is delivered, that may be actioned in real-time to improve performance. The Health Online Model Ensemble Serving Framework (HOLMES), described originally in 2020, is a well-documented example of a Deep Learning real-time ensemble system used to provide clinical decision support to physicians. HOLMES is used in high acuity, complex settings such as the Intensive Care Unit (ICU) where there is an abundance of constantly changing data that is available to the system¹². HOLMES is unique in that it is capable of assembling a unique predictive model for each patient in an ICU setting by drawing from more than one model. Real-time predictions provide a high value-add opportunity to guide real-time decisions within established workflows. In the analysis that follows, we apply the RCI framework to HOLMES to examine its relevance to VBC, with the appreciation that HOLMES is restricted to a high acuity setting with an abundance of dynamic real-time data.

1.1 Relevance:

Potential to mitigate key VBC challenges
 HOLMES directly addresses the challenge of delayed performance data by providing real-time predictions. Traditional VBC performance reporting suffers from data lags, whereas

Demonstrates quantifiable improvement in quality of care and cost efficiency
 According to the study conducted by Shenda Hong and others (2020), HOLMES demonstrated high prediction accuracy (above 95%) in a pediatric cardiac ICU setting¹³. This level of accuracy suggests

¹¹ IBM. What is deep learning? IBM. Published February 26, 2025. <https://www.ibm.com/think/topics/deep-learning>

¹² Hong S, Xie Y. HOLMES: Health OnLine Model Ensemble Serving for deep learning models in intensive care units. In: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. Association for Computing Machinery; <https://doi.org/10.1145/3394486.3403212>

¹³ Hong S, Xu Y, Khare A, Priambada S, Maher K, Aljiffry A, Sun J, Tumanov A. HOLMES: Health OnLine Model Ensemble Serving for deep learning models in intensive care units. In: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. Association for Computing Machinery; <https://doi.org/10.1145/3394486.3403212>

HOLMES and similar models work on streaming data, offering immediate insights into clinical risks and treatment effectiveness. Real time accurate predictions offer physicians the opportunity to intervene immediately, thereby influencing the clinical trajectory.

significant potential for improvement. HOLMES has been directly associated with improved quality in the ICU setting as a predictor of mortality. Further, the ability of HOLMES to predict length of stay and step-down readiness thereby facilitating earlier transitions of care, has direct bearing on cost efficiencies. This cost advantage is magnified from the continuously online HOLMES system, compared to the traditional offline “batching” method, where readiness for transition from the high cost ICU setting may be anticipated and proactively prepared for without unwarranted delays.

1.2 Context:

Portability based on resources

HOLMES is designed specifically for the ICU and uses a complex infrastructure (graphic processing units and robust networks), implying the system may not be easily transferable to other health care settings. More widely, adoption of Deep Learning models in resource-constrained environment, may struggle due to the required computational demands potentially making these models less transferable between differing environments. While the conceptual principles of real-time risk prediction should be broadly applicable across a wide range of settings, adapting models that are trained on highly specialized data (e.g., pediatric cardiac ICU) to other clinical scenarios will require significant retraining and validation.

Usability

The inherent complexity of Deep Learning models such as HOLMES can pose a challenge for usability. System users would require substantive training to understand the basis of the predictions, how to interpret the results, and how to apply decision logic to appropriately integrate model insights into real-time clinical workflows.

1.3 Workflow Integration:

Digital workflow integration

Deep Learning models such as HOLMES, that are used for clinical risk prediction, are highly dependent on integration with EHR systems to access real-time patient data, which would in turn require uniform data formats such as HL7. Integration is crucial for real-time data access, as these models require continuous, high-frequency data feeds. Further, clinical decision support is a key function of HOLMES, amplifying the need for workflow integration. Deep Learning models may require significant customization to integrate within existing workflows due to the aggregation of data from multiple sources in the clinical environment. Such customizations may require modifications to data capture systems, data

Transparency of the application

While the complexity of HOLMES provides the advantage of accuracy, the same complexity risks the perception of the system as a poorly understood 'black box'. This lack of transparency poses a barrier to adoption since clinicians must understand the basis of the prediction to trust and rely on the model for making critical clinical decisions. The use of “Explainable AI” (XAI) techniques can make these models more transparent by providing insights into what factors are influencing the model's predictions, although these techniques are also complex to implement. An additional consideration involves the model's treatment of patient data that requires careful attention to anonymization, secure data

management infrastructure, and the structure of the EHR itself.

storage, shared liability, and access controls that are compliant with privacy regulations such as HIPAA.

Summary:

Deep Learning models, exemplified by HOLMES, offer a significant step forward in addressing the real-time data needs of VBC. The code for HOLMES has been made available by its developer in the GitHub. While HOLMES was analyzed as a Use Case, several online ensemble applications are available for monitoring patients in high acuity settings, for example, Epic System's ICU Module, GE Healthcare's CARESCAPE Monitoring Systems and Philips IntelliVue Guardian solution. These proprietary applications similar to HOLMES provide methods for real time monitoring in ICU settings, each with a similar significant requirement of investment in infrastructure and training. Future development is required to enhance the transparency of Deep Learning models to realize the full opportunity for real-time decision making, and to assess the potential of these models in other care settings

2. Natural Language Processing to reduce Administrative burden

The time required for transcribing and structuring the patient encounter into an EHR standard is a key underlying driver of the administrative burden experienced by physicians when consulting patients¹⁴. The Natural Language Processing (NLP) model offers the clinician the unique opportunity to hand over the transcribing function to the AI model, thereby significantly easing the administrative burden facing health care professionals within clinical workflows¹⁵. NLP is being used within clinical consultations as a chatbot for language translation, and encounter summarization. Essentially, NLP involves the real-time processing of unstructured data associated with clinical conversations using Ambient Clinical Voice (ACV), a technology whereby the health care professional records the patient encounter. At the conclusion of the encounter, the recording is transformed into a draft clinical note that can be added to the EHR¹⁶. Given the transformative promise of the technology, it is not unexpected to observe the recent proliferation of ACV vendors in the AI health care ecosystem¹⁷.

2.1 Relevance:

Potential to mitigate key VBC challenges

Early studies indicate that ambient clinical voice solutions may reduce clinical documentation time by up to 50%¹⁸. In addition to the obvious advantage of providing the physician more time to enhance the quality of the patient interaction, ACV technologies offer the opportunity for more structured and comprehensive note-taking with the potential to influence the delivery of clinical care itself. Practically, writing support provided by ACV aims to enhance the quality of clinical

Demonstrates quantifiable improvement in quality of care and cost efficiency gains

Administrative burden has been definitively associated with professional burnout, contributing as much as 60%. NLP has been demonstrated to improve professional burnout resulting from administrative burden²⁰, in turn resulting in improved quality of care. A small study noted an improvement in patient experience in ACV enabled encounters²¹. We hypothesise that ACV enabled patient encounters that offer the physician more time

¹⁴ Leong HY, Gao YF, Shuai J, Zhang Y, Pamuksuz U. Efficient fine-tuning of large language models for automated medical documentation. arXiv. Published September 2024. <https://arxiv.org/abs/2409.09324>

¹⁵ Alafari F, Driss M, Cherif A. Advances in natural language processing for healthcare: a comprehensive review of techniques, applications, and future directions. Compute Methods Programs Biomed Update. <https://www.sciencedirect.com/science/article/abs/pii/S1574013725000024>

¹⁶ Blum K. All ears: what to know about ambient clinical listening. Association of Health Care Journalists. Published March 18, 2024.

<https://healthjournalism.org/blog/2024/03/all-ears-what-to-know-about-ambient-clinical-listening/>

¹⁷ Healthcare IT Today. Ambient clinical voice companies. Healthcare IT Today. Published February 28, 2025. <https://www.healthcareittoday.com/ambient-clinical-voice-companies/>

¹⁸ Nuance Communications. Nuance and SCIENTIA Puerto Rico expand Dragon Medical One access to improve patient outcomes and experiences. PR Newswire. Published June 30, 2022.

²⁰ Lawyer R. Combating clinician burnout with AI: a 2025 vision for smarter healthcare workflows. Unite.AI. Published January 24, 2025.

<https://www.unite.ai/combating-clinician-burnout-with-ai-a-2025-vision-for-smarter-healthcare-workflows/>

²¹ Owens LM, Wilda JJ, Grifka R, Westendorp J, Fletcher JJ. Effect of ambient voice technology, natural language processing, and artificial intelligence on the patient-physician relationship. Mayo Clinic Proc Digit Health. 2024. <https://europepmc.org/article/MED/38834180>

notes using functions such as auto-completion and auto-structuring¹⁹. However, unstructured and unlabelled data will have to be closely monitored to guarantee accuracy. There is inherent heuristic value in how physicians record and use their notes to guide the diagnostic process, where the note is a reflection of the logic and thought process a physician applies to clinical evaluation. It is important for ACV technologies to preserve, or enhance this value in the patient encounter. An additional complexity involves settings of care where human scribes already fulfill the clinical documentation role. It will be interesting to observe to what extent ACV technologies disrupt this role.

with the patient, will increase patient trust, thereby improving patient adherence to the treatment plan. This goal is particularly relevant in primary care for patients with behaviourally modifiable chronic conditions where non-adherence can be very costly. Diagnostic safety is an additional quality of care dimension that stands to gain from the improved quality of the clinical encounter, and subsequent improved referral and follow-up plan. More, and higher quality, time spent with the physician, should mean more time both for the diagnostic process and to influence patient behaviors, resulting in improved outcomes of care. NLP usage may lead to an increase in the volume of data captured as observed in the case of 'Nuance Dragon Medical One' where a 167% increase in the documentation captured has been described²². The high accuracy performance of ACV, combined with the increase in the volume of data captured has the potential in the long term to generate novel data sets. Such data could be used to enhance the predictive power of machine learning models that operate as continuous learning systems, thereby influencing clinical decision-making and further improving quality outcomes.

2.2 Context:

Portability based on resources

NLP-based applications can be used in most health care settings where EHRs are used. It is likely that ACV in its current design will have limited application in settings that employ paper-based systems. The plethora of NLP applications and EHR systems imply that interoperability will likely be challenging as adoption increases, thereby increasing procurement costs and adoption time²³. ACV systems represent a significant shift from the traditional patient-physician interaction and offer the immediate the opportunity to reduce administrative burden. More so, these technologies offer the transformative opportunity to fundamentally re-design the clinical encounter in ways that intentionally improve quality outcomes, cost efficiencies and patient experience.

Usability

NLP applications are user friendly requiring minimal user training. The technologies are sensitive to language nuances including linguistic variations and dialects. Users may be trained in the effective usage of the key technical terms. Transcription errors are expected in the early phase of adoption²⁴, involving substitution errors, omission errors or terminology errors²⁵.

¹⁹ Aramaki E, Wakamiya S, Yada S, Nakamura Y. Natural language processing: from bedside to everywhere. *Yearb Med Inform.* 2022. <https://www.thieme-connect.com/products/ejournals/abstract/10.1055/s-0042-1742510>

²² Nuance Communications. AI-powered speech recognition eases clinical documentation burden. Nuance. Published 2025. https://www.nuance.com/content/dam/nuance/en_us/collateral/healthcare/case-study/cs-dragon-medical-one-embedded-in-epic-allina-health-en-us.pdf

²³ Velupillai S, Suominen H, Liakata M, et al. Using clinical natural language processing for health outcomes research: overview and actionable suggestions for future advances. *J Biomed Inform.* 2018. <https://pubmed.ncbi.nlm.nih.gov/30368002/>

²⁴ Zhou L, Blackley SV, Kowalski L, et al. Analysis of errors in dictated clinical documents assisted by speech recognition software and professional transcriptionists. *JAMA Netw Open.* 2018. <https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2687052>

²⁵ Medical Transcription Service Company. What are the serious errors that can occur in medical transcription. Medical Transcription Service Company. Published December 20, 2021. <https://www.medicaltranscriptionservicecompany.com/blog/what-are-serious-errors-that-occur-in-medical-transcription/>

1.3 Workflow Integration:

Digital workflow integration

Most ACV systems operate in parallel to the EHR, requiring a manual intervention to add the ACV note to the EHR. This bifurcated manifestation of ACV is likely a result of the nascency of the technology and can be expected to change as the technology matures and adoption increases. An additional challenge involves the inherent nature of the data itself, pertaining to privacy and utility. Legal frameworks necessitate that data be de-identified or anonymized before being applied to NLP models. This limits access to certain demographic identifiers (for example geolocation data) with the potential to compromise decision-making within workflows.

Transparency of the model

NLP applications require explicit user consent and permissions for processing and storage. This requirement, coupled with immediate access to data and the ability for corrections at any stage, increases transparency of the models. Advantageous to health care uses, NLP models are already commonly used across various mobile applications for day-to-day tasks, with the potential to mainstream the technology and increase acceptance. Familiarity with NLP applications outside of health care, associated with reliability and value, stands to augment the trust that patients will place in the technology when deployed within health care settings.

Summary:

NLP has emerged as a powerful intervention to address the challenges surrounding the administrative burden facing health care professionals. The ease of use and transparency associated with the first NLP models, combined with the clinician's and patient's familiarity with the technology in uses outside of health care, offer the potential for rapid adoption. A key limiting factor involves the limited integration of the application with existing EHR's, likely a result of the early stage of evolution of the technology. Rigorous studies are required to demonstrate how reductions in administrative tasks, impact quality outcomes and costs.

3. Large Language Models to improve EHR interoperability

In 2013, the Office of the National Coordinator for Health Information Technology (ONC) began tracking interoperable exchange across hospitals in the US. A decade later, in 2023, 70% of non-federal acute care hospitals reported all four domains of interoperable exchange (send, receive, find, and integrate), compared to 46% in 2018. The admirable progress, however, masks substantive variation across hospital size and location (urban/rural). The FHIR (Fast Healthcare Interoperability Resources) specification developed within the international standard, HL7, has streamlined EHR interoperability enabling seamless data flows between EHR applications, standalone applications and patient-generated health data (PGHD)²⁶.

Despite the adoption of FHIR, full interoperability has still not been achieved due to inconsistent implementation. This lag in interoperability perpetuates the well described fragmentation in service delivery in the nation. Mitigating interoperability is costly, providing an opportunity for Large Language Models (LLM) to add value. This can be accomplished by the capability that LLMs possess to provide a common layer across IT applications, that are usually not compatible with each other, to interact across. LLMs enhance quality of care by synthesising patient data from incompatible EHR applications, providing clinical decision support system with higher precision and reduced administrative burden. One notable example is MedLM, developed by Google²⁷ as a 'family of foundation models fine-tuned for healthcare industry use cases'. In the context of enhancing interoperability, Accenture has built on MedLM to interpret structured and unstructured data from multiple sources and automate the manual process.

3.1 Relevance:

Potential to mitigate key VBC challenges

LLMs can directly support VBC goals and improve quality of care by increasing access to accurate patient data, the basis for clinic decision-making. Access to wide variety of data (structured and unstructured) leads to better accuracy of AI models. A study conducted to analyse the effects of AI in EHR systems through Health Information Exchanges showed that the AI models produced positive predictive values from 83.70 % to 94.10 % and negative predictive values between 94.10 % and 99.10 %²⁸, demonstrating the potential for LLMs to augment HIEs and interoperability.

There are several applications for LLMs within the context of VBC. By processing unstructured health data, LLM's are able to extract previously unknown (unidentified) risk factors thereby improving clinical risk predictions. When analysing variation across practices, LLM's are able to "standardize" data between providers, thereby rendering patient records interoperable through novel mechanisms. Additionally, LLMs have the potential to generate longitudinal patient summaries by aggregating data

Demonstrates quantifiable improvement in quality of care and cost efficiency

LLMs have been applied to unstructured data representing complex medical narratives (including patient histories and physician notes) to generate meaningful information. These models utilize advanced natural language processing methods such as contextual embedding - where words or phrases are interpreted based on their meaning in specific contexts to standardize and map the extracted data to FHIR specifications. According to the study done by Yikuan Li the LLMs have demonstrated a greater than 90% accuracy rate in converting clinical unstructured texts into FHIR compatible resources²⁹. The high accuracy may be translated into time savings combined with a reduction in manual errors during transcriptions. Any gains to interoperability from the use of LLMs will undoubtedly improve quality of care due to narrowing fragmentation gaps.

²⁶ HealthIT.gov. Health Level 7 (HL7) Fast Healthcare Interoperability Resources (FHIR). HealthIT.gov. <https://www.healthit.gov/topic/standards-technology/standards/fhir>

²⁷ Matias Y, Gupta A. Introducing MedLM for the healthcare industry. Google Cloud Blog. Published December 14, 2023. <https://cloud.google.com/blog/topics/healthcare-life-sciences/introducing-medlm-for-the-healthcare-industry>

²⁸ Borna S, Maniaci MJ, Haider CR, et al. Artificial intelligence models in health information exchange: a systematic review of clinical implications. J Med Internet Res. 2023. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10531020/>

²⁹ Li Y, Wang H, Yerebakan HZ, Shinagawa Y, Luo Y. Enhancing health data interoperability with large language models: a FHIR study. medRxiv. Preprint posted online October 21, 2023. <https://doi.org/10.1101/2023.10.17.23297028>

dispersed across multiple providers through both historical and future encounters.

3.2 Context:

Portability based on resources

The deployment of LLMs requires access to advanced computational resources which includes Graphics Processing Units (for model training, evaluation and inference), cloud storage (for handling large scale data sets including image files) and scalable computing options. Hospitals with access to this infrastructure, with personnel proficient in AI usage will be able to harness the value of LLM applications. In contrast, smaller hospitals and those in rural settings will likely face challenges in adopting LLM models. This dynamic is paradoxical in that entities that currently lag in interoperability exchange, risk falling further behind as LLMs, operationalized at organizational level, advance.

Usability

There is limited empirical evidence pertaining to the ease of usability of LLMs to improve EHR operability. While it is reasonable to assume, based on the early use of LLMs, that these models will not be associated with significant personnel training requirements, interoperability is complex and may present unique challenges that reduce the ease of use.

3.3 Workflow Integration:

Digital workflow integration

LLM applications can be easily integrated with existing data sources - both structured (lab results and EHRs) and unstructured (clinical notes). The regulatory frameworks that will govern the use of LLMs pertaining to personal health data will likely pay equal attention to patient consent and health professional validation.

Transparency of the model

LLM models are highly complex and are likely to present a significant challenge with adhering to desired transparency levels. The models yield varying results for even slightly different inputs which erodes user trust. Commonly prevailing model models such as GPT-4 and MedLM function as 'black boxes' with low levels of explicability when compared with traditional rule-based decision support systems. The models are exposed to bias and fairness considerations, and are dependent on the nature of the data sets on which training occurs, the latter itself being an obfuscated process. Recent developments in the field of "Explainable AI" (XAI) attempt to address this gap. XAI incorporates rule-based hybrid models which combine the flexibility of LLMs with hard-coded clinical rules³⁰. The confidence scores associated with the XAI compatible AI models demonstrate how reliable a model's predictions are, rather than strictly measuring accuracy. The evolving fairness-aware AI techniques and bias audits can be used to analyze a LLM's performance, compared to an expected standard, in all stages of model building, from data gathering, labelling, pre-processing and results summarization.

³⁰ IBM. What is explainable AI? IBM. Published March 29, 2023. <https://www.ibm.com/think/topics/explainable-ai>

Summary:

LLMs provide new avenues for enhancing EHR interoperability. The seamless integration of data facilitated by LLMs has potential to improve the quality of clinical decisions. However, potentially high costs, limited transparency, and evolving usage rules will be limiting factors for adoption³¹. Explainable AI offers the opportunity to mitigate some of these challenges as LLMs are advanced as productivity enablers within health care ecosystems. Also, LLM access to the sensitive patient data across various systems requires explicit patient consent and compliance with HIPPA and state specific privacy laws to ensure its ethical use and data security.

³¹ Eastwood B. The evolution of LLMs in healthcare. HealthTech Magazine. Published July 2024. <https://healthtechmagazine.net/article/2024/07/future-llms-in-healthcare-clinical-use-cases-perfcon>

4. Remote Patient Monitoring to enhance Quality improvement

Quality Improvement (QI) strategies, focused on improving processes and outcomes, are dependent on an organization's ability to measure performance, identify gaps, and design, implement and evaluate strategies. Technological and data management capabilities are central to the infrastructure required for the effective implementation of QI. Traditional sources of data are skewed to what is easily accessible within health care systems, and as described previously, these are usually associated with time lags. Remote Patient Monitoring (RPM) through wearables provides a technological solution to enhancing QI efforts by collecting unique patient level data. AI powered RPM analyses vast amounts of data from the wearables using advanced time series forecasting. Long Short Term Memory (LSTM) is a type of neural network that uses sequential data points collected over a period of time to identify patterns. ML algorithms are used to detect abnormal patterns that are used for predictive modelling, for example, anomaly detection ML algorithms using decision trees in RPM may detect sudden drops or spikes in signals such as glucose levels or oxygen rates. The opportunity to continuously monitor a patient for disease progression or worsening, and act timeously, represents a dramatic shift in how care is delivered, with significant potential to improve quality and costs. To monitor disease progression, health data may be continuously transmitted through the RPM mechanism and tracked by the patient's health provider. For example, Continuous Glucose Monitors (CGM) tracks blood sugar fluctuations and reports them to the clinician. A study highlighted by the American Diabetes Association found that RPM integrated with CGMs, to measure blood sugar levels over several months, reduced Haemoglobin A1C levels by 1% in Type 2 diabetes patients³². Similar RPM powered applications in monitoring disease progression exist for heart failure facilitating clinical evaluation over an extended period without a physical visit to the clinic. Real-time data from the wearables enables timely interventions augmented by predictive analytics. Additionally, the same real-time feedback to the patient has the potential to reinforce treatment adherence goals.

4.1 Relevance:

Potential to mitigate key VBC challenges

Readmissions remain one of the key focus areas for QI strategies globally. It is estimated that approximately \$17 billion out of \$26 billion spent annually on Medicare re-admissions are avoidable (KFF). Gaps in post-discharge monitoring have been consistently identified as the main contributing factor³³ (CMS). To disincentive sub-optimal discharge planning, the Hospital Readmission Reduction Program (HRRP) launched in 2012, penalizes hospitals with higher than expected readmissions³⁴. There is undoubtedly significant potential for wearables that permit continuous real-time monitoring to be used as solutions to mitigate the risk of readmissions. Hospital at Home programs are natural platforms off which to pivot the application of wearables in this manner.

Demonstrates quantifiable improvement in patient outcomes and cost efficiency

Clinical data collected from wearable devices provides the opportunity to contribute new sources of patient data to transform clinical health care data sets. AI powered RPM, enabled through wearables, helps in generating early warnings signals in patients suffering from chronic diseases or offers benefit in post-surgical monitoring, with direct implications for improving cost and quality. Preventing unnecessary admissions and readmissions can be expected to be a key goal of RPM. The UK's National Health Services (NHS) executed a pilot program in 2019, to analyse the vital health data of patients in real-time. Algorithms constantly monitored the patient's vital signs using a Wi-Fi enabled armband) for warning signals with alerts triggered to patients and physicians when clinical thresholds were breached. The study concluded that the intervention resulted in

³² American Diabetes Association. Breakthrough studies on automated insulin delivery and CGM for type 2 diabetes unveiled at ADA Scientific Sessions. American Diabetes Association. Published June 24, 2024. <https://diabetes.org/newsroom/press-releases/breakthrough-studies-automated-insulin-delivery-and-cgm-type-2-diabetes>

³³ Centers for Medicare & Medicaid Services. Findings from recent CMS research on Medicare by Niall Brennan. Centers for Medicare & Medicaid Services.

³⁴ Centers for Medicare & Medicaid Services. Hospital Readmissions Reduction Program (HRRP). CMS.gov. Published September 10, 2024. <https://www.cms.gov/medicare/quality/value-based-programs/hospital-readmissions>

A study conducted by Pavithra Lakshman and others compared traditional early warning scores implemented through 3 to 4 spot checks daily to remote patient monitoring with an 'automated remote early warning system (R-EWS)'³⁵. The R-EWS real time models showed high sensitivity (97.37%) - ability to identify patients at risk of deterioration of health status. The average time between the initial alert and the actual deterioration was a minimum 18 hours for remote patient monitoring systems which showed both increased alert frequency and higher proportion of critical alerts. RPMs offer significant potential to disrupt disease progression with concomitant benefits for quality and costs.

4.2 Context:

Portability based on resources

Wearable devices such as smart watches and bio sensors are designed to transmit large amounts of real-time collected clinical data to remote servers for real-time continuous monitoring. This requires significant IT capabilities and investment accompanied by a significant re-design of clinical workflows involving the re-allocation of clinical personnel. The upfront investment for such transformative change is high, and will need to be balanced by the long-term benefits, which are likely to be substantive.

Usability of the application

Wearables and other RPM solutions do not require a higher level of technical skills from the user as these are mostly plug-and-pay mode devices. However, the cost of procuring the wearable device, digital literacy, uninterrupted connections to clinical personnel, and proximity to physical clinical services, are important considerations.

4.3 Workflow Integration:

Digital workflow integration

RPM applications which are FHIR compliant are designed to seamlessly exchange data between various healthcare applications. Epic's App Orchard provides a marketplace for applications that integrate with its applications. Similarly, Cerner's open developer experience encourages applications that integrate with its' applications. Generally, applications that adhere to FHIR standards can be seamlessly integrated with the EHR and relevant digital workflows.

Transparency of the model

AI models involving RPM have a low level of explicability and their "black box" nature may present challenges with adoption. It is crucial that patients hold high levels of trust in RPM mechanisms given the dependency on patients to directly handle and engage with the technology.

Summary:

RPM applications provide promising avenues for transforming QI strategies. However, in doing, so they present new challenges pertaining to algorithmic explicability and the requirement for high upfront investments. Empowering patients with the skills to act as a partner in managing care is an equally important consideration.

³⁵ Lakshman P, Gopal PT, Khurdi S. Effectiveness of remote patient monitoring equipped with an early warning system in tertiary care hospital wards: retrospective cohort study. J Med Internet Res. 2025. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC11780298/>

Acknowledgements

We offer our gratitude to the many people who supported our work and gave generously of their time and advice, and for whose guidance and mentorship we are grateful for:

Faculty Advisor: Joseph P. Newhouse, esteemed John D. MacArthur Professor of Health Policy and Management at Harvard University;

Senior Fellows and the leadership team at the Mossavar-Rahmani Center for Business and Government, Harvard Kennedy School: Professor Richard Zeckhauser, Professor John Haig, Dan Murphy, Susan Gill;

Peers, colleagues and friends at the College of Health Solutions, Arizona State University; and, Harvard graduate students who participated in our study group sessions and who inspire us with their commitment to improving health care systems around the world.

References:

1. Centers for Medicare & Medicaid Services (CMS). Premier Hospital Quality Incentive Demonstration. Accessed May 28, 2025. <https://www.cms.gov/medicare/quality/initiatives/hospital-quality-initiative/premier-quality-incentive-demonstration>
2. Centers for Medicare & Medicaid Services (CMS). How MIPS Eligibility Is Determined. Published September 2024. Accessed April 11, 2025. <https://qpp.cms.gov/mips/how-eligibility-is-determined>
3. Centers for Medicare & Medicaid Services (CMS). Traditional MIPS Scoring Guide for the 2024 Performance Year. Published June 2024. Accessed April 11, 2025. <https://qpp-cm-prod-content.s3.amazonaws.com/uploads/2875/Traditional-MIPS-Scoring-Guide-for-the-2024-Performance-Year.pdf>
4. Health Care Payment Learning and Action Network (HCPLAN). APM Measurement, Progress of Alternative Payment Models. 2024 Methodology and Results Report
5. Sylvia M. Burwell. Setting Value-based payment goals - HHS efforts to improve US Health Care. The New England Journal of Medicine. 2019; DOI: 10.1056/NEJMp1500445
6. Rachel M. Werner, Ezekiel J. Emanuel et al. The Future of Value-Based Payment: A Road Map to 2030. Penn LDI Leonard Davis Institute of Health Economics. 2021
7. American Medical Association (AMA). Letter to Chiquita Brooks-LaSure, CMS Administrator, Regarding MIPS Cost Performance Category. October 27, 2023. Accessed April 11, 2025. <https://searchlf.ama-assn.org/letter/documentDownload?uri=%2Funstructured%2Fbinary%2Fletter%2FLETTERS%2Ffcsls.zip%2F2023-10-27-Letter-to-Brooks-LaSure-re-MIPS-Cost-Performance-Category-v2.pdf>
8. Hockenberry JM. Time and Financial Costs for Physician Practices Participating in the Medicare Merit-based Incentive Payment System: A Qualitative Study. JAMA Health Forum. Published February 10, 2021. Accessed April 11, 2025. <https://jamanetwork.com/journals/jama-health-forum/fullarticle/2779947>
9. Turbow S, Hollberg JR, Ali MK. Electronic health record interoperability: How did we get here and how do we move forward? JAMA Netw Open. Published 2021. Accessed April 29, 2025
10. Centers for Medicare & Medicaid Services (CMS). Artificial Intelligence Demo Days. Published November 5, 2024. <https://www.cms.gov/digital-service/artificial-intelligence-demo-days>

11. IBM. What is deep learning? IBM. Published February 26, 2025. <https://www.ibm.com/think/topics/deep-learning>
12. Hong S, Xie Y. HOLMES: Health OnLine Model Ensemble Serving for deep learning models in intensive care units. In: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. Association for Computing Machinery; <https://doi.org/10.1145/3394486.3403212>
13. Hong S, Xu Y, Khare A, Priambada S, Maher K, Aljiffry A, Sun J, Tumanov A. HOLMES: Health OnLine Model Ensemble Serving for deep learning models in intensive care units. In: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. Association for Computing Machinery; <https://doi.org/10.1145/3394486.3403212>
14. Leong HY, Gao YF, Shuai J, Zhang Y, Pamuksuz U. Efficient fine-tuning of large language models for automated medical documentation. arXiv. Published September 2024. <https://arxiv.org/abs/2409.09324>
15. Alafari F, Driss M, Cherif A. Advances in natural language processing for healthcare: a comprehensive review of techniques, applications, and future directions. Compute Methods Programs Biomed Update. <https://www.sciencedirect.com/science/article/abs/pii/S1574013725000024>
16. Blum K. All ears: what to know about ambient clinical listening. Association of Health Care Journalists. Published March 18, 2024. <https://healthjournalism.org/blog/2024/03/all-ears-what-to-know-about-ambient-clinical-listening/>
17. Healthcare IT Today. Ambient clinical voice companies. Healthcare IT Today. Published February 28, 2025. <https://www.healthcareittoday.com/ambient-clinical-voice-companies/>
18. Nuance Communications. Nuance and SCIENTIA Puerto Rico expand Dragon Medical One access to improve patient outcomes and experiences. PR Newswire. Published June 30, 2022.
19. Aramaki E, Wakamiya S, Yada S, Nakamura Y. Natural language processing: from bedside to everywhere. Yearb Med Inform. 2022. <https://www.thieme-connect.com/products/ejournals/abstract/10.1055/s-0042-1742510>
20. Lawyer R. Combating clinician burnout with AI: a 2025 vision for smarter healthcare workflows. Unite.AI. Published January 24, 2025. <https://www.unite.ai/combating-clinician-burnout-with-ai-a-2025-vision-for-smarter-healthcare-workflows/>
21. Owens LM, Wilda JJ, Grifka R, Westendorp J, Fletcher JJ. Effect of ambient voice technology, natural language processing, and artificial intelligence on the patient-physician relationship. Mayo Clinic Proc Digit Health. 2024. <https://europepmc.org/article/MED/38834180>
22. Nuance Communications. AI-powered speech recognition eases clinical documentation burden. Nuance. Published 2025. https://www.nuance.com/content/dam/nuance/en_us/collateral/healthcare/case-study/cs-dragon-medical-one-embedded-in-epic-allina-health-en-us.pdf
23. Velupillai S, Suominen H, Liakata M, et al. Using clinical natural language processing for health outcomes research: overview and actionable suggestions for future advances. J Biomed Inform. 2018. <https://pubmed.ncbi.nlm.nih.gov/30368002/>

24. Zhou L, Blackley SV, Kowalski L, et al. Analysis of errors in dictated clinical documents assisted by speech recognition software and professional transcriptionists. JAMA Netw Open. 2018. <https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2687052>
25. Medical Transcription Service Company. What are the serious errors that can occur in medical transcription. Medical Transcription Service Company. Published December 20, 2021. <https://www.medicaltranscriptionservicecompany.com/blog/what-are-serious-errors-that-occur-in-medical-transcription/>
26. HealthIT.gov. Health Level 7 (HL7) Fast Healthcare Interoperability Resources (FHIR). HealthIT.gov.. <https://www.healthit.gov/topic/standards-technology/standards/fhir>
27. Matias Y, Gupta A. Introducing MedLM for the healthcare industry. Google Cloud Blog. Published December 14, 2023. <https://cloud.google.com/blog/topics/healthcare-life-sciences/introducing-medlm-for-the-healthcare-industry>
28. Borna S, Maniaci MJ, Haider CR, et al. Artificial intelligence models in health information exchange: a systematic review of clinical implications. J Med Internet Res. 2023. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10531020/>
29. Li Y, Wang H, Yerebakan HZ, Shinagawa Y, Luo Y. Enhancing health data interoperability with large language models: a FHIR study. medRxiv. Preprint posted online October 21, 2023. <https://doi.org/10.1101/2023.10.17.23297028>
30. IBM. What is explainable AI? IBM. Published March 29, 2023. <https://www.ibm.com/think/topics/explainable-ai>
31. Eastwood B. The evolution of LLMs in healthcare. HealthTech Magazine. Published July 2024. <https://healthtechmagazine.net/article/2024/07/future-llms-in-healthcare-clinical-use-cases-perfcon>
32. American Diabetes Association. Breakthrough studies on automated insulin delivery and CGM for type 2 diabetes unveiled at ADA Scientific Sessions. American Diabetes Association. Published June 24, 2024. <https://diabetes.org/newsroom/press-releases/breakthrough-studies-automated-insulin-delivery-and-cgm-type-2-diabetes>
33. Centers for Medicare & Medicaid Services. Findings from recent CMS research on Medicare by Niall Brennan. Centers for Medicare & Medicaid Services.
34. Centers for Medicare & Medicaid Services. Hospital Readmissions Reduction Program (HRRP). CMS.gov. Published September 10, 2024. <https://www.cms.gov/medicare/quality/value-based-programs/hospital-readmissions>
35. Lakshman P, Gopal PT, Khurdi S. Effectiveness of remote patient monitoring equipped with an early warning system in tertiary care hospital wards: retrospective cohort study. J Med Internet Res. 2025. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC11780298/>