

Impact

Duke
Institute
for Health
Innovation

Ideas that challenge. Innovation that inspires.

Volume 26



**Quality,
Access, Equity,
and Impact**

**Catalyzing
Transformative
Innovation and
Research**

**Inventions,
Thought
Partners and
Generative AI**



Contents

04	Introduction	
06	Impact at a Glance	
08	DIHI is on the Move	
09	RFA Projects' Impact	
28	Spotlights	
36	Experience with DIHI	
42	AI Governance & Ethics	
51	Expanding the Portfolio	
61	Publications & Collaborators	



The Duke Institute for Health Innovation

Impact Report Volume 26



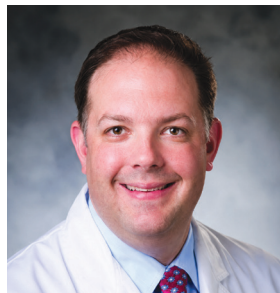
About DIHI

The Duke Institute for Health Innovation (DIHI) catalyzes transformative innovations in health and healthcare through implementation of high-impact innovations, leadership development, and cultivation of a community of entrepreneurship. We bring innovative solutions to the most pressing challenges in health and healthcare through multidisciplinary teamwork across Duke University and Duke Health and by fostering collaborations with national and international thought leaders.

INTRODUCTION



Suresh Balu, MS, MBA



Jeff Ferranti, MD, MS

Letter From the Executive Director and the CDO

Over the past year, changing policies and trends have reshaped our health care innovation strategy. As health systems everywhere have continued to face workforce shortages, financial challenges and shifting regulatory patterns, so too have they benefited from the unprecedented opportunities afforded by responsible use of GenAI, and leaps in the integration of digital technologies to better meet patient needs. Amid these dynamic changes, Duke Health continues to prioritize innovation across the entire system keeping both patients and care teams at the center and staying true to its mission.

This annual report outlines key solutions our team implemented, including an initiative that improves integration of health-related social needs into electronic health records (EHRs). By leveraging large language models, we make it easier for care teams to access and organize information about social drivers, enabling them to quickly connect patients with community resources.

Another key project aims to automate oncology patient access in Wake County, a rapidly growing area with 1.3 million residents. DIHI is improving cancer care access by automating new patient data collection, which previously required manual review of charts, imaging, and lab reports. The AI system streamlines intake so staff can quickly match patients with appropriate providers, reducing workload and expediting care. This method, currently used for oncology, is scalable to other specialties to further improve access and patient experience.

This year, we launched Scout, an advanced AI platform that empowers care teams to seamlessly manage intricate EHR tasks using natural language. Scout swiftly searches and synthesizes patient data with full provenance, dramatically boosting efficiency, easing cognitive demands, and driving down care costs while

also accelerating research capabilities. As Scout is deployed across service lines, Duke is licensing the platform to other health systems, exponentially expanding its pioneering influence and transforming healthcare nationwide.

This report further highlights Duke Health's national impact through the expansion of the Health AI Partnership. We have established a practice network and technical assistance program to support FQHCs and other providers in underserved communities. The first cohort included five health systems, and we are exploring pathways to address the AI digital divide in collaboration with North Carolina's Office of Rural Health.

As we look ahead, digital health represents more than just a trend; it is a fundamental evolution in healthcare delivery. We are transitioning from a centralized, reactive, and episodic model toward a distributed, proactive, and continuous care system, that centers entirely on the individual. DIHI is uniquely positioned to lead this transformation, and we are committed to supporting the mandate to deliver the highest quality care to a growing number of patients.

Finally, innovation is inherently complex, demanding a rare blend of creative risk-taking and disciplined execution. The DIHI team exemplifies the convergence of diverse knowledge, from programming expertise and product management to a relentless, can-do attitude. But we know innovation doesn't happen in isolation. Our success is only possible thanks to the robust ecosystem of collaboration with partners like Duke Health Technology Solutions, the Population Health Management Office, Duke Heart Center, and many more. To all our collaborators, thank you! Your support is the foundation of every accomplishment in this report.



Impact at a Glance

RFA Projects' Impact

AI-Assisted Surgery Prior Authorization

p. 9

Orthopedic teams lose significant time preparing documentation for prior authorization reviews. DIHI developed an AI solution that evaluates surgical notes against ten high-value authorization quality criteria. Across 40 complex notes, the tool achieved 98.5% accuracy; over 100 prospective surgical plans validated by five Duke Orthopedic surgeons, accuracy reached 99.8%. The system helps schedulers identify documentation gaps before submission, reducing denials and administrative rework.

Duke Health Policy Chat

p. 12

Duke Health staff previously navigated more than 4,000 policy documents spread across disparate formats, leading to delays and compliance risk. DIHI built a Retrieval Augmented Generation (RAG)-powered chat assistant that searches all Duke Health policies in real time and delivers grounded, accurate answers. Deployed systemwide to PIN stations across DUHS, the tool-now rebranded Duke Health Policy Chat and integral to Duke's institutional policy platform-has answered hundreds of queries monthly, with documented adoption across perioperative and broader clinical teams.

Clinical Education Exploration Dashboard

p. 13

Medical educators lack a timely, systematic method to assess whether students encounter the full breadth of clinical experiences required for training. DIHI built an LLM-powered Clinical Education Dashboard in Tableau that analyzes student-authored Epic notes against 40 required and 43 recommended diagnoses. During the 2024-25 academic year, 123 second-year students authored more than 11,200 notes documenting nearly 150,000 diagnostic encounters. Clerkship directors may now use the dashboard in real time to close exposure gaps and guide curricular decisions.

AI for Health-Related Social Needs

p.19

Social determinants of health are often buried in unstructured clinical notes, limiting the ability of care transition teams to act at scale. DIHI's solution uses generative AI via Microsoft Azure to synthesize HRSN data at discharge, automatically generating a structured HRSN Discharge Summary Note in Maestro Care. Since going live November 1, 2025, the tool has generated notes for 13,491 hospitalizations. Between Nov 1, 2025 and May 1, 2026, clinician-graded quality scores averaged above 4.75 out of 5 across all domains, with estimated annual time savings of 975 hours for the DukeWELL Care Transitions Team.

CTICU AI Handover System

p. 15

Shift-change handovers in the Cardiothoracic ICU are a critical-and cognitively demanding point of failure. DIHI developed an AI handover system that automatically retrieves relevant notes from Epic and generates standardized summaries delivered to incoming physicians via email or Microsoft Teams. Validated across 8,924 outputs by two critical care physicians, the system achieved 99% accuracy on required structural elements and 98.4% on manual fact checks. Deployment is expanding to additional critical care units, with Epic integration under development.

Medication Prior Authorization

p.17

Prior authorization for high-cost medications is a manual, time-consuming process that drives staff burnout, delays care, and exposes Duke Health to financial risk when denials go uncontested. DIHI's AI tool retrieves real-time patient data from Maestro Care, evaluates it against payer-specific criteria using generative AI, and drafts complete submission and appeal letters. Validated at 100% accuracy across 30 Ocrevus cases and 99% across 31 Botox cases, the tool went live with the pre-service verification team in January 2026, with Phase 2 expansion to GLP-1 inhibitors and additional therapies underway.

Oncology Access Assistant

p.22

Newly diagnosed cancer patients need fast access to the right oncologist - but manual chart review by New Patient Coordinators limits throughput and strains a growing access center. DIHI's Oncology Access Assistant automatically retrieves clinical notes, imaging, pathology, and lab results (including from outside health systems via Epic's Care Everywhere), then applies scheduling rules via LLM to recommend the correct provider and timeframe in 14 seconds - more than 50 times faster than manual review. Live since March 3, 2026, the solution supported 77 new patient triages in its first three weeks, saving an estimated 15.7 staff hours.

Kidney Transplant Matching

p.25

When a donor kidney becomes available for transplant, the surgeons must rapidly decide if the organ matches a specific recipient. While every offer must go through various filters to ensure a baseline level of organ quality, the final organ acceptance decision requires synthesizing donor characteristics, recipient medical complexity, and contextual factors. To support this decision-making process under time pressure, our team developed a machine learning model that leverages donor and recipient data to predict the risk of short- and long-term kidney dysfunction following kidney transplantation. This tool aims to improve match quality, standardize decision-making, and improve patient outcomes.



Impact at a Glance

Spotlights

Scout: Your Thinking Partner

p. 28

A blinded randomized crossover trial demonstrated a 40% reduction in cognitive load and 40% reduction in EHR task time across seven use cases - establishing the first rigorous evidence base for AI-assisted chart review as a tool for both clinician well-being and patient safety.

LLM Ops: Standardizing Clinical AI Pipelines

p. 30

DIHI's engineering infrastructure for deploying and monitoring clinical LLM pipelines at scale, anchored by EHRx, DIHI's FHIR-native successor to Monorail.

What We've Built - and How to Keep It

p. 32

DIHI's impact came from an integrated, compounding, deeply embedded team and infrastructure that took a decade to build. We grow more valuable with every project we produce. This article is about understanding that system, what threatens it, and what sustaining it requires.

AI Governance & Ethics

HAIP Programmatic Update

p. 42

HAIP gives healthcare organizations frameworks, tools, and peer networks for responsible AI adoption. This article documents growing uptake of HAIP's Key Decision Points and deepening engagement with partner health systems navigating AI governance.

HAIP Practice Network Experience

p. 44

The first full Practice Network cycle brought health systems together to advance AI governance maturity. Participating organizations shared implementation experiences and evaluation frameworks, informing HAIP's tools and policy guidance on safe AI adoption at scale.

Vendor Disclosures for AI

p. 47

A new HAIP framework structures vendor disclosure across algorithm performance, infrastructure, safety, and governance - mapped to HAIP's 8 Key Decision Points and designed for both large academic medical centers and resource-limited organizations.

Solution Facts Label for Generative AI

p. 49

Extending DIHI's Model Facts Label to generative AI, whose open-ended outputs require a different disclosure approach. The Solution Facts Label provides tiered visibility for regulators, institutional administrators, clinical staff, and patients.

Expanding The Portfolio

Scaling ML Deployment Infrastructure

p. 51

DIHI has systematically worked to build reusable infrastructure for deploying machine learning models at scale. Drawing on lessons from multiple production deployments, the article covers model monitoring, version management, drift detection, and the organizational processes required to sustain model performance over time. Our FHIR-based EHRx product provides the technical foundation for much of this scaling work.

Inbasket Messaging with LLMs

p. 53

Physician inbasket overload is one of the most measurable contributors to clinician burnout. We describe DIHI's LLM-based inbasket messaging solution, which drafts and categorizes patient messages to reduce the manual burden on care teams. The solution is in active use, with ongoing measurement of time savings and clinician satisfaction as the team prepares for broader deployment.

Scaling Adult Deterioration Detection

p. 54

Sepsis Watch and related early warning models have long operated within DUH. In 2025-2026, DIHI scaled adult deterioration detection to DRAH and DRH, expanding the reach to thousands of additional patients. Since go-live in Raleigh in Jan 2026, the system has generated 190 notifications across 73 unique hospitalizations. Duke Regional is in active validation, with workflows adapted to each site.

Surgery CADEN: AI for Surgical Complexity

p. 58

The DIHI-Surgery Innovation Lab team built an AI solution to support the assessment and communication of surgical complexity across surgical services. The Clinical Condition Ascertainment & Detection Engine (CADEN) integrates structured and unstructured data to characterize case complexity, supporting scheduling, resource allocation, and documentation quality. Early validation data show strong concordance with expert assessments, and the team is actively pursuing integration with surgical workflows.



DIHI is on the Move

Scout Use 9/2025-05/2026

>800
Scout Users

>30
Specialties

>35,000
Scout Tasks

RFA Projects' Impact

AI for Health-Related Social Needs

Generates structured HRSN Discharge Summary Notes in Maestro Care from unstructured clinical text. Live Nov 1, 2025. Clinician quality scores averaged above 4.75/5. Estimated 975 annual staff hours saved.

Clinical Education Dashboard

11,200+ student notes, 150,000 diagnostic encounters, 123 students tracked in real time.

Duke Health Policy Chat

RAG assistant across 4000+ policies. Now used more than 150 times per day.

Oncology Access Assistant

Triage new cancer patients using EHR + Care Everywhere data. Recommends provider and timeframe in 14 seconds - 50x faster than manual review. Live March 3, 2026; 77 triages in first three weeks.

AI-Assisted Surgery Prior Authorization

Evaluates surgical notes against 10 quality criteria. Validated across 100 prospective plans by 5 Duke Orthopedic surgeons.

CTICU AI Handover System

Standardized shift summaries via Teams or email. Validated across 8,924 outputs.

Medication Prior Auth

Duke AI evaluates payer criteria and drafts submission and appeal letters. Live Jan 2026.

Transplant Matching

Predicting delayed graft function to improve Kidney Transplant Matching.

Spotlights

Scout: Your Thinking Partner

Blinded RCT across 7 use cases establishes first rigorous evidence base for AI-assisted chart review and synthesis platform.

LLM Ops & EHRx

DIHI's FHIR-native EHRx standardizes clinical AI pipelines at scale. Protectable IP with licensing potential.

What We've Built - and How to Keep It

A decade of compounding infrastructure. Understanding the system, what threatens it, and what sustaining it requires.

Expanding The Portfolio

Scaling Adult Deterioration Detection

Adult Deterioration went live at Duke University Hospital (DUH) in June 2025 and Duke Raleigh Hospital (DRAH) in January 2026. At DUH, 4517 notifications (12 per day) have been sent on 1815 hospitalized patients; at DRAH, 326 notifications (2 per day) have been sent on 146 hospitalized patients.

Surgery CADEN

Improves clinical documentation integrity using structured + unstructured data. Strong concordance with expert assessments.

RFA Project Impact

Using AI to Reduce the Burden of Prior Authorizations

Alison Toth, MD
John Bonini, BS
Erik Carvalho, PT, DPT, MHA
Will Knechtle, MBA, MPH
Mike Revoir, BS
Amy Podraza, MBA
Suresh Balu, MS, MBA



Problem

Surgery pre-authorization is increasingly complex. This complexity slows the efficient provision of timely patient care. This challenge is dually driven by payers' increasing documentation requirements and providers' insufficient documentation. According to data from the Patient Revenue Management Office (PRMO), the Department of Orthopedics saw a 15% increase in surgery authorization requests from early 2023 to late 2024. Manual processes to resolve these challenges have become labor intensive, especially for the Orthopedic surgeons, administrative staff, and PRMO staff.

The peer-to-peer consultation workload (time an orthopedic surgeon must talk directly with an insurance company's doctor to discuss the medical reasons for a proposed surgery) has increased to several hours per week in 2025, when it was once 1-2 hours per month in 2023. These additional hours increase Duke's valuable human capital costs while often delaying or removing revenue.

Solution

We implemented an AI solution to streamline the surgery pre-authorization process in the Department of Orthopedics. This AI tool automatically identifies and outlines

the necessary documentation criteria for surgery authorizations, reducing the likelihood of missing critical elements, minimizing delays in patient care, preventing administrative errors, and mitigating revenue loss. It streamlines surgical scheduler work and frees up healthcare providers to concentrate on direct patient care.

Our first step was to identify prior authorization criteria for spine and arthroplasty plans and assessments submitted to insurers. The Duke PRMO team reviewed prior authorization denials over five years and determined the most common needs across Spine, Hip, Knee, and other joint surgery notes for generic prior authorization (from most health insurers).

Today, when an orthopedic surgeon, revenue manager, or surgery-scheduler selects a signed note for AI application review, it checks the note for completion of ten criteria. The user can quickly read the AI solution's completed checklist and related reasoning per criterion. Then, a surgical scheduler or care practitioner can improve the note before it is sent out to an insurance company. The tool is not meant to make treatment decisions, but rather, evaluate how those treatment decisions were documented.

Process Results

We partnered with orthopedic surgeons, lead administrators, and PRMO analysts to create a gold standard prior authorization result for each of 40 notes with orthopedic surgery plans. We developed ten high-value authorization quality checkpoints and tested how well our AI solution performed compared to expert humans' gold standard. In this set of 40 complex surgical assessments and convoluted notes, the average result score was 98.5%. After this retrospective step, five Duke Orthopedic surgeons reviewed 100 plans' AI checklist completion over a period of 3 months. When scaled to 100, that is 1000 questions, the accuracy was 99.8%. Orthopedic surgeons have reported their satisfaction after a decline in documentation rework and fewer peer-to-peer calls with payor representatives.

Impact

The impact of this solution is substantial. It can streamline surgery authorization and enhance technology integration, resulting in more efficient, cost-saving processes. This will improve patient care, reduce cancellations, and increase operational efficiency. The collaborative effort involving DIHI, PRMO, Duke Compliance Office, DHIP, and the Department of Orthopedics promises to streamline healthcare processes, leading to a transformative impact across Duke University Health System.

Criterion	Accuracy
1. Clear Name and Specification of Surgical Procedure	100%
2. Explicit Documentation of the Procedure Status	97.5%
3. Documentation of Physical Therapy (PT) Visits	100%
4. Documentation of PT Visit Frequency/Volume/Timeline	100%
5. Documentation of Steroid Injection Use	95.0%
6. Documentation of Non-Surgical Alternative Treatments	97.5%
7. Explicitly Documented the Prior MRI or CT	100%
8. Explicitly Documented an X-Ray	100%
9. Note Specifies the Exact Date of Imaging	95.0%
10. Explicit Documentation of Tobacco Smoking Status	100%

Table 1: AI accuracy across 10 prior authorization quality criteria (n=100 complex notes).



See a real example AI solution output below, de-identified

AI Solution Double Checks the Note Selection, Providing Summary and Reasoning:

This note is a definitive orthopedic surgical assessment because it includes a concrete surgical plan (C3-7 posterior decompression with laminoplasties), documents patient consent and understanding, and reviews relevant clinical findings and imaging. It is not a peripheral plan, injection note, follow-up, or early meeting; it is the latest note that establishes a clear intent to proceed with surgery and would be sent to a prior-authorizing payor for review ^[1].

Task 02: Prior Authorization Criteria Assessment

Summary Output Format

- Surgical procedure documentation: ☒
Rationale: Anatomical site and approach are clearly documented ^[1].
- Surgical service status documented: ☒
Rationale: No explicit documentation of inpatient, outpatient, ambulatory, or same-day status in this note.
- PT visit history documented: ☒
Rationale: PT attendance is documented ^[1].
- PT visits' frequency OR volume OR timeline documented: ☒
Rationale: No explicit documentation of PT frequency, volume, or timeline in this note.
- Steroid injection administered OR attempted (not just considered): ☒
Rationale: No documentation of steroid injection for cervical spine in this note.
- Alternative non-surgical non-steroidal non-PT-visit treatments: ☒
Rationale: Water aerobics documented as alternative therapy ^[1].
- Prior MRI or CT explicitly documented: ☒
Rationale: Prior MRI is explicitly documented ^[1].
- Prior or very-soon X-ray explicitly documented: ☒
Rationale: No explicit documentation of X-ray in this note.
- IMAGING DATE: ☒
Rationale: MRI date is explicitly documented ^[1].
- Tobacco smoking status OR history documented: ☒
Rationale: No documentation of tobacco smoking status or history in this note.

Task 03: Checklist Items Failed – Additional Facts and Contents Needed

Criteria Failed	What Should Be Added	Source & Citation
Surgical service status documented	Explicitly state whether the surgery will be inpatient, outpatient, ambulatory, or same day. Example: "This will be performed as an inpatient procedure."	The preoperative evaluation note (PASS) documents the location as "DUKE NORTH OR" but does not specify inpatient/outpatient/ambulatory status ^[2] . No other note provides explicit service status.
PT visits' frequency/volume/timeline	Add details on PT duration, frequency, or timeline (e.g., "Patient completed 8 weeks of PT, 2x/week").	The MM/DD/YYYY (obfuscated for publication) spine surgery evaluation note documents "She had physical therapy and water aerobics 6 months" ^[3] . The MM/DD/YYYY cardiology note states "[Pronoun] has attended PT for about a year" ^[4] . These details should be added to the main surgical assessment note.
Steroid injection administered/attempted	Add documentation if any cervical epidural steroid injection (ESI) or similar was given or attempted for the cervical spine.	No evidence of cervical steroid injection. Prior hip injections are documented, but not relevant for cervical surgery ^[5] .
Prior or very-soon X-ray explicitly documented	Add documentation of cervical spine X-ray, including date or relative time.	The MM/DD/YYYY PASS note documents "Cervical spine X-ray MM/DD/YY " with findings ^[2] . This should be referenced in the main surgical assessment note.
Tobacco smoking status OR history documented	Add explicit documentation of tobacco smoking status/history.	The MM/DD/YYYY PASS note documents "Smoking status: Never" ^[2] . The MM/DD/YYYY cardiology note also documents "Smoking status: Never" ^[4] . These should be added to the main surgical assessment note.

Additional Advice for Bulletproof Prior Authorization:

- Add explicit service status (inpatient/outpatient/ambulatory/same day).
- Reference PT duration and frequency (e.g., "PT for 6 months, 2x/week").
- Reference cervical spine X-ray and its date.
- Add explicit tobacco smoking status.
- If steroid injection for cervical spine was not performed, document rationale for omission (e.g., contraindication, not indicated).

Limitations:

There is no documentation of cervical steroid injection, nor explicit service status in the main surgical assessment note. All other required elements can be supported by referencing other notes.

Duke Health Policy Chat

Matt Gardner, BS
Erin Hamtil, RN
Samantha Fiske, RN
Katy Clark, RN
Will Knechtle, MBA, MPH
Michael Kent, MD
Suresh Balu, MS, MBA

The Problem

Previously, Duke Health providers struggled to obtain timely access to critical policy content, with over 4,000 documents scattered across different formats such as PDFs and Word files. This lack of centralized, easily searchable information often resulted in delays when seeking answers to policy questions, increasing the risk of care interruptions, compliance lapses, and inconsistent practices among staff; all of which could negatively affect patient outcomes.

The need for a more efficient method to quickly access and interpret policy information was evident. To address these barriers, the DIHI 2024 RFA initiative, "AI-Powered Knowledge Management System and Training Module," developed an AI-powered assistant to deliver accurate and prompt responses to Duke Health policy questions.

The Solution

To address the challenge of timely access to policy content, DIHI developed a chat assistant powered by a large language model (LLM) using an approach called Retrieval Augmented Generation (RAG). With thousands of policy documents scattered across different formats, this approach was especially effective for our situation.

The RAG approach enabled the LLM to answer questions by first searching through all the policy documents to find the most relevant sections, or "snippets," that matched what a user was asking. These relevant snippets from Duke Health's actual policies were then provided to the LLM, allowing it to generate accurate and helpful responses. This meant that staff could efficiently get answers grounded in the latest policy content, instead of struggling to search through the documents themselves.

Adoption and Full-Scale Deployment

The DIHI Policy Chat was first implemented with a small clinical team focused on perioperative policy content. During the pilot phase in summer 2024, the tool consistently provided relevant and reliable support, leading to growing interest across DUHS.

For broader adoption, all Duke Health policies were integrated via automated synchronization with the Duke Health Policy Center. Deployed to all Duke Health PIN stations in fall 2024, it was rebranded Duke Health Policy Chat and is now integral to Duke's institutional policy platform.

Impact and Utilization

Monthly utilization has grown from ~10 queries at launch to over 2,600/week by April 2026, reflecting system-wide adoption.

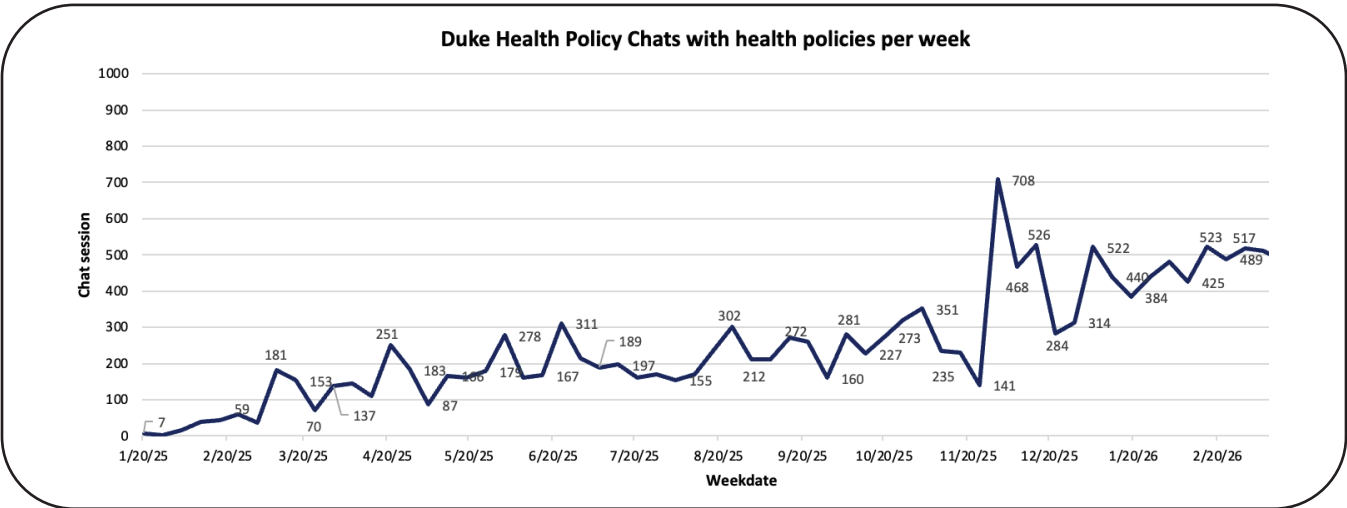


Figure 1: Duke Health Policy Chat weekly query volume (Jan 2025 – Apr 2026). Growth from pilot to full-system deployment is visible.

Clinical Education Exploration Dashboard Using LLMs to Improve Medical Education

Will Knechtle, MBA, MPH
 Matt Gardner, BS
 Joel Boggan, MD
 Genie McPeck Hinz, MD, MS
 Suresh Balu, MS, MBA

The Problem

Medical education relies heavily on clinical experiences to bridge the gap between classroom learning and real-world patient care. However, traditional methods for evaluating these experiences are limited. Most schools depend on retrospective, manual reporting from students and preceptors, which is subject to recall bias, inconsistency, and significant administrative burden.

Despite the vast amount of data generated in Electronic Health Records (EHRs), student-authored clinical notes are rarely leveraged to systematically assess the breadth and depth of trainee exposure to key diagnoses and procedures. This gap makes it difficult for educators to ensure that all students encounter the conditions and skills deemed essential for a comprehensive medical education.

Solution

To address these challenges, we developed a novel method that harnesses Large Language Models (LLMs) to analyze and summarize medical student-authored clinical notes from Epic. This approach integrates narrative note content with encounter metadata to provide a comprehensive, real-time view of trainee clinical experiences.

The Clinical Education Dashboard (CED) is an interactive Tableau-based tool that visualizes and quantifies student exposure to "Blue Recs" — a curated list of 40 required and 43 recommended diagnoses and procedures prioritized by the Duke University School of Medicine.

During the 2024–25 academic year, 123 second-year medical students authored over 11,200 notes, documenting nearly 150,000 Blue Recs diagnoses. The dashboard provides detailed breakdowns of exposure to high-priority conditions such as cardiovascular disease, congestive heart failure, diabetes, liver disease, and pneumonia.

Process & Academic Impact

By automating the extraction and analysis of clinical experiences, the dashboard significantly reduces administrative burden on both students and faculty. Educators can now proactively identify students underexposed to certain conditions and intervene early. The dashboard also opens new avenues for educational research, enabling correlation of clinical exposure patterns with learning outcomes, board exam performance, and residency placement.

NEXT PAGE →



Academic Impact

Academically, the Clinical Education Dashboard represents a pioneering application of artificial intelligence in medical education. By leveraging LLMs to extract and map clinical experiences, the project provides a scalable, reproducible method for aligning real-world training with curricular goals. Early results demonstrate strong alignment between student-documented diagnoses and institutional educational targets.

The dashboard also opens new avenues for educational research, enabling correlation of clinical exposure patterns with learning outcomes, board exam performance, and residency placement. The project has already generated scholarly output, including presentations and abstracts, and positions Duke as a leader in the integration of AI and medical education.

Conclusion

The Clinical Education Dashboard project envisions how advanced AI methods can transform the measurement and management of clinical learning. By providing actionable, data-driven insights into trainee experiences, the dashboard empowers educators and students to align training with meaningful, measurable patient care exposure, ultimately improving the quality and efficiency of medical education at Duke and other academic medical centers.

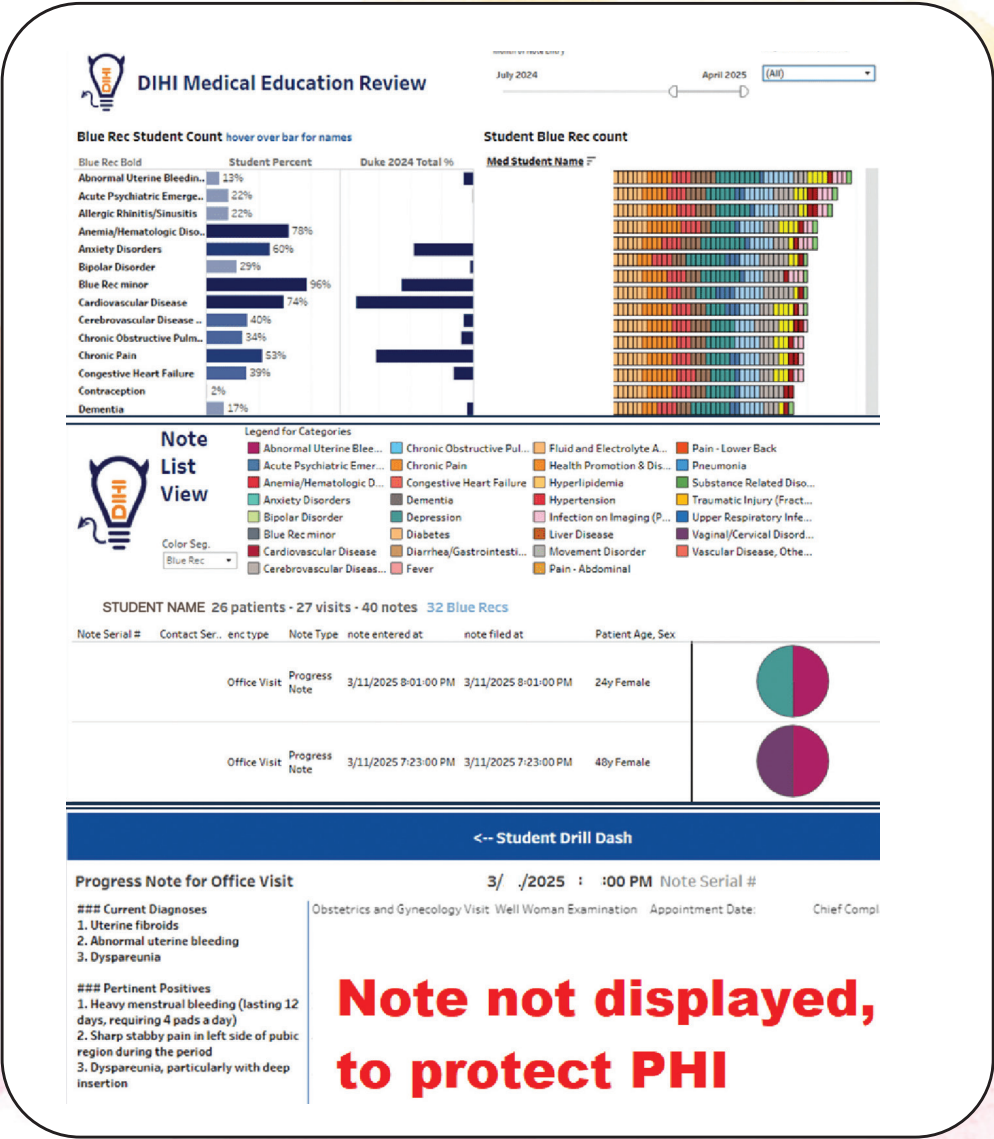


Figure 1: Clinical Education Dashboard interface. 123 students tracked across 40 required and 43 recommended diagnoses, with 11,200+ notes and 150,000 documented encounters. PHI removed.

Implementing an AI-Powered CTICU Handover Summarization Tool

Sachin Mehta, MD
Tareq Aljurf, MD
Mike Revoir, BS
Will Knechtle, MBA, MPH
Suresh Balu, MS, MBA

The Problem

The Cardiothoracic Intensive Care Unit (CTICU) at Duke University Hospital is a high-volume, high-acuity environment where patient care demands are both complex and dynamic. Each day, attending physicians must perform critical shift turnovers, communicating essential patient information to the incoming team. Currently, this handover process relies on manual summarization of patient status, typically via email or verbal communication. These summaries are highly variable in content, length, and quality, and lack consistent semantic stratification. In other words, the urgency, severity, and prioritization of clinical concerns are not reliably conveyed. The manual nature of this process is further complicated by physician fatigue, especially at the end of long shifts, and by the sheer volume of unstructured data embedded in clinical notes. As a result, vital information may be omitted, miscommunicated, or lost in translation, potentially impacting patient safety and care continuity.

The affected demographic included 15 attending CTICU physicians who rotate through shifts and are responsible for the care of critically ill patients, including those undergoing heart and lung transplants, mechanical circulatory support, and complex cardiac or thoracic surgeries. The lack of a standardized, efficient, and comprehensive handover process not only increases cognitive burden on clinicians but also introduces risk to patient outcomes due to inconsistent communication.

Solution

To address these challenges, the CTICU Handover project proposed the development and implementation of an AI-powered assistant designed to automate and standardize the CTICU handover process. Leveraging advanced large language modeling (LLM) and generative AI, the tool

extracts and summarizes CTICU daily event logs, assessment, and plan sections within the electronic health record (EHR) handoff documentation spaces. This material often contains nuanced clinical reasoning, intent, and concern that are not captured in structured data fields.

The AI solution was trained on data from the CTICU's attending physicians, ensuring that its output aligns with clinical expectations and workflow needs. More than 97 cases with 23 measures over four runs, meaning 8,924 outputs, were validated by two critical care physicians and checked by two medical students. The DIHI Handover system generated relevant, accurate, and comprehensive shift summaries for each patient, incorporating predefined key components to ensure consistency. 99% of the outputs contained the true/false checks required, 99.6% contained the desired text order and volume, and 98.4% met the manual fact accuracy checks. For context, in the world of biostatistics and generative models in the year 2025, this is exceptional.

In application, at the end of each shift, the CTICU Handover AI automatically retrieves relevant notes and generates a standardized summary for each patient. The summary is then securely delivered to the incoming attending physician, either via email or a Duke Microsoft Team channel.

The phased implementation included staff training and full-scale deployment, with ongoing monitoring and iterative improvement based on user input. Full-scale deployment has included interest from several other critical care units and derivatives for other hospital services. The Microsoft Team channel has not been an agreeable interface for physicians, so we are pursuing alternative interfaces and Epic integrations to enable full scale use.



Process Impact

The introduction of the DIHI AI Handover Solution is poised to transform the Duke University Health Systems’ hospital unit shift-handover process in several key ways:

- 1. **Standardization and Consistency** through strictly prompted, physician-designed structure. This reduces variability in content and quality, mitigating the risk of omissions or miscommunication.
- 2. **Efficiency and Clinician Well-being:** Automation alleviates the cognitive and administrative burden on care practitioners, particularly at the end of demanding shifts. This not only saves time but also reduces the impact of fatigue, enabling clinicians to focus on direct patient care and decision-making.
- 3. **Improved Documentation Quality:** Knowing that daily notes will be used for automated handover summaries incentivizes physicians to maintain high-quality, comprehensive documentation, further enhancing the reliability of the process.
- 4. **Feedback and Continuous Improvement:** The system incorporated many early feedback loops and was created through a deep partnership entirely between Duke engineers, data scientists, physicians, and nurses. This ensured relevant and effectiveness. Pre- and post-deployment surveys will measure satisfaction and completeness of handovers, guiding ongoing optimization.

Potential Business Impact

- 1. **Patient Safety and Outcomes:** By reducing communication errors and ensuring that critical information is consistently conveyed during handover.
- 2. **Operational Efficiency:** Measurable time savings for physicians can free up resources for other clinical duties and potentially reduce overtime costs. It may also enhance staff satisfaction and retention.
- 3. **Scalability and System-wide Adoption:** While initially focused on the CTICU, the DIHI Handover platform serves as a template for AI-driven clinical note summarization across other units and specialties. This amplifies impact on clinician efficiency and care coordination.
- 4. **Competitive Advantage and Innovation Leadership:** By pioneering the use of generative AI in clinical handovers, Duke Health positions itself at the forefront of healthcare innovation. This not only supports recruitment and retention of top clinical talent but also strengthens the institution’s reputation as a leader in digital health transformation.

- 5. **Cost Savings and ROI:** Over time, reductions in adverse events, improved workflow efficiency, and enhanced documentation quality may translate into tangible cost savings. The initial investment in development and implementation is modest relative to the potential long-term benefits in quality, safety, and operational performance.

Conclusion

The DIHI Handover represents a strategic investment in AI-driven clinical workflow optimization, addressing a critical need in the CTICU for standardized, efficient, and reliable shift handovers. By leveraging advanced LLMs and generative AI, the tool not only improves the quality and consistency of communication among attending physicians but also laid groundwork for broader adoption of AI solutions in Duke hospitals. For example, we worked immediately with the Surgical Intensive Care Unit (SICU) to replicate our methods and we received approval to pilot the handover there. Furthermore, the 2026 request for applications resulted in two funded Duke Health DIHI projects that scale this service across our health system.

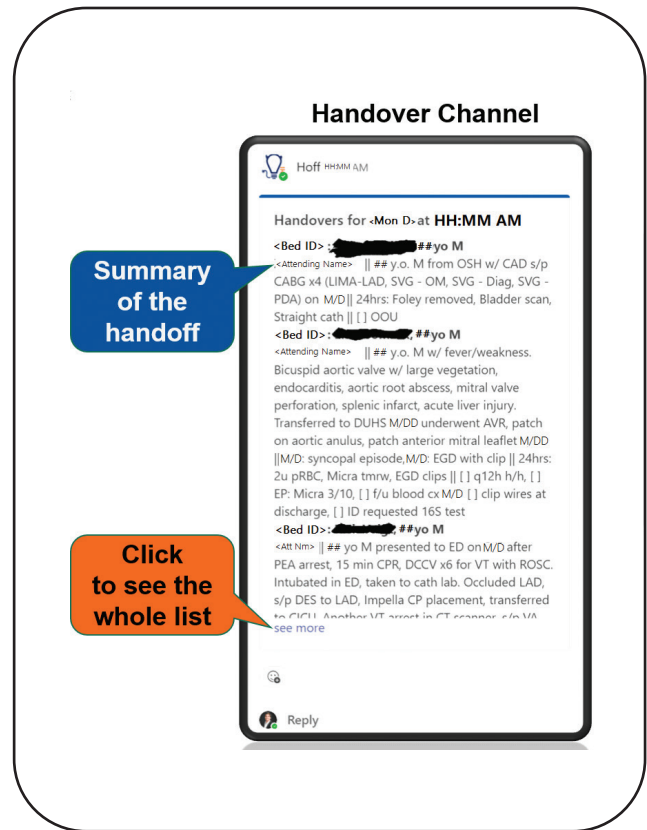


Figure 1: CTICU AI Handover tool delivered via Microsoft Teams (PHI removed). Placed here from the adjacent CTICU article on the following pages.

Using AI to Reduce the Burden of Medication Prior Authorization

Joanna Kipnes, MD, MS
Deborah Kaye, MD, MS
Bobette Haley
Anand Chowdhury, MD, MMCI
Danyang Zhuo, PhD
Will Ratliff, MBA
Mike Revoir, MS
Nikkie Dutta, MS
Anika Mohan
Marshall Nichols, MS
Suresh Balu, MS, MBA



Project Brief

Prior authorization (PA) requests create a significant administrative burden at Duke Health, contributing to staff and physician burnout, delays in care, and increased costs for the health system. To address this challenge, our team developed an AI-based solution that automates chart review to surface relevant patient data and prepare prior-authorization requests or denial appeals. The tool is currently being implemented and evaluated in real-time, with metrics including time to complete PA requests, rate of medication denials, appeal turnaround time, staff administrative time, and overall impact on patient care and outcomes.

Problem

Prior authorization (PA) is utilized by insurers to assess whether a proposed medication or medical service is medically necessary,¹ requiring approval from a patient's health plan before a patient can access certain health-care goods and services, including medications, surgeries, radiology, and laboratory testing. If PA is not obtained, insurers can deny coverage of the medication or medical service. However, the current process is manual and inefficient as providers are often unaware of payer-specific criteria, multiple staff must compile and submit documentation, and relevant information is not consistently highlighted or complete. Additionally, high-cost medications often must be paid for up-front by the health system at Duke, with reimbursement contingent on prior authorization approval. When requests are denied, this can lead to significant financial losses and delays in patient care.

Solution

To address these challenges, Joanna Kipnes, Deborah Kaye, the PRMO team, and the Duke Institute for Health Innovation developed an AI platform to automate chart review and draft prior authorization submissions and denial appeals. Phase 1 of the project focuses on medications with high average monthly volume, denials, and treatment costs, and include Ocrevus, Botox for Migraine, Pluvicto, and Darzalex Faspro.

For each medication, we created a Medical Prior Authorization Assistant (MPAA) tool that retrieves patient data from Maestro Care (e.g., provider notes, flow sheets, labs, orders, medications, and ICD codes) in real-time and combines it with payer-specific prior authorization and medical necessity criteria. See Figure 1 for the MPAA clinical workflow for Ocrevus. On the solution's user interface, users enter the patient's MRN. The current date auto-populates to support authorization submissions for patients' upcoming medication infusions. It can be adjusted to support denial appeal workflows. The user then selects "search" to pull in the relevant chart information and "Evaluate the patient for Ocrevus Prior Authorization". The patient's chart is then evaluated by a Large Language Model (LLM) to determine whether the patient meets criteria for the requested medication therapy. When all authorization criteria are met, the solution generates a documentation checklist for user to follow to select the relevant records in the release-of-information (ROI) workflow. It also drafts a cover letter that incorporates supporting clinical information for submission or appeal. If the criteria is not met, the solution tells the user to stop and lists the criteria that were met as well as those that were not met.

To validate the solution, we evaluated retrospective prior authorization cases for all four medications, including approvals, denials, and supporting data. From these cases, we created "ground truth" files that included the authorization status and relevant supporting data. These ground truth cases were then compared to outputs generated by the LLM. Each individual criterion was marked as a pass if the LLM output matched the ground truth. If a criterion generated by the LLM did not match ground truth, we reviewed the case, refined the prompt and augmented the raw source data as needed, and re-tested to confirm a subsequent match between LLM and ground truth.

Impact

The evaluations for Ocrevus, Botox for Migraine, and Pluvicto have all been completed, with scores of 100%, 99%, and 98% respectively (see Table 1). These scores reflect the LLM’s accuracy in matching the output of the ground truth files. Notably, the LLM correctly assessed the authorization status for all 83 cases across the three medications, with rare permutations from ground truth related to supporting documentation (7 errors total out of 970 criteria assessed). The Medical Prior Authorization Assistant tool for Ocrevus went live on January 12, 2026 with the Medical Necessity & Denials Team. The Medical Prior Authorization Assistant tool for Botox went live on February 23rd with the PSR Team and financial care counseling teams.

Next Steps

We plan to go live in May 2026 with the Outpatient Pre-Service Verification (PSV) teams for Ocrevus, Pluvicto, and Darzalex Faspro. In Phase 2, we are currently refining the prompt and evaluating the tool to authorize GLP 1 inhibitors, Pembrolizumab, and Romidepsin and go live with these medications in Summer 2026. The long-term goal is to continue testing, refining, and scaling the solution across Duke Health to support prior authorization for a broader range of services, including radiology, procedures, laboratory testing, and durable medical equipment.

References:

[1] Gupta R, Fein J, Newhouse J P, Schwartz A L. Comparison of prior authorization across insurers: cross sectional evidence from Medicare Advantage BMJ 2024; 384 :e077797 doi:10.1136/bmj-2023-077797

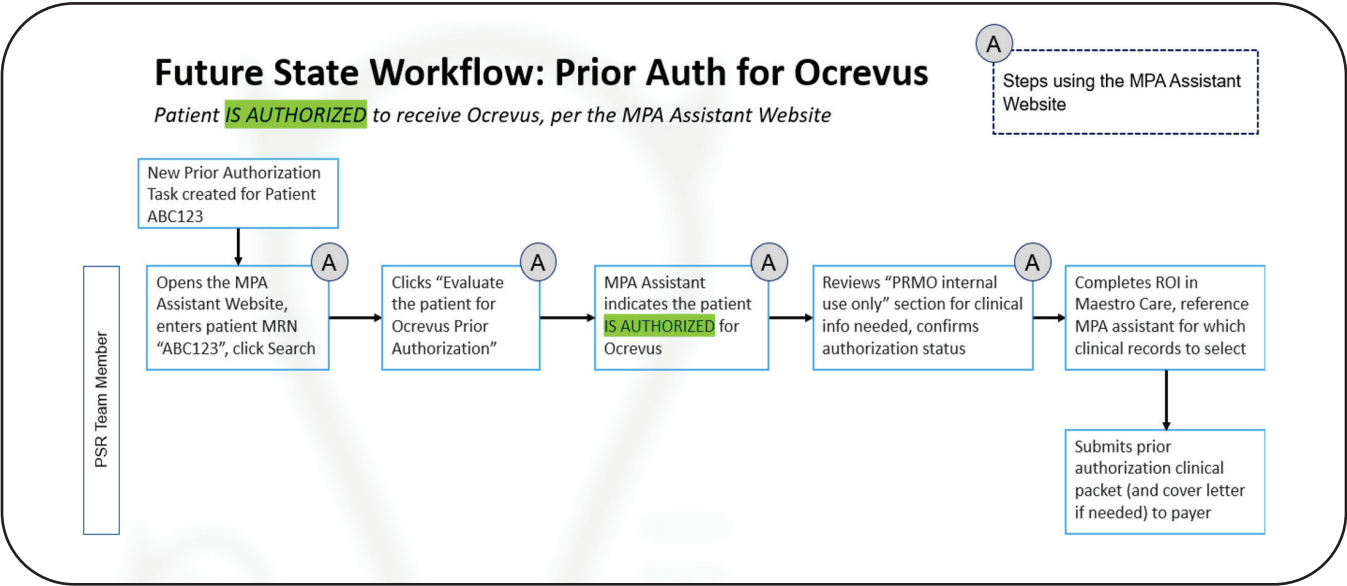


Figure 1: MPAA clinical workflow for Ocrevus

Medication	Criteria/Case	Cases	Criteria Passed
Ocrevus	12	30	360/360 (100%)
Botox for Migraine	14	31	430/434 (99%)
Pluvicto	8	22	173/176 (98%)

Table 1: Evaluation Results for Ocrevus, Botox for Migraine, and Pluvicto

AI for Improving Health Related Social Needs Care for Duke Patients

Genie McPeck Hinz, MD
Susan Spratt, MD
Tara Kinard, DNP, MBA
Will Ratliff, MBA
Noah Prizant, MD
Mike Revoir, BS
Tres Brown, BS
Bradley Hintze, PhD
Michael Gao, PhD
Suresh Balu, MS, MBA

Project Brief

Duke provides high quality health care to its patients. It also provides support for its patients' health-related social needs (HRSNs). Due to the high volume of patients with HRSNs and the chart review work required to support an intervention for a patient, it is challenging to provide comprehensive care for all Duke patients' HRSNs. To address this challenge, we use artificial intelligence to automatically create a HRSN Discharge Summary Note for all hospitalized patients, available in the electronic health record on the day after they are discharged. Beginning in November 2025, our clinical teams have relied on this note to improve the speed and completeness of their preparation for a HRSN intervention on a patient.

Problem

Unmet health-related social needs (HRSNs) worsen health outcomes, lead to hospital re-admissions, and increase healthcare costs. Current screening approaches to identify HRSNs are inconsistent, incomplete, and often ineffective at connecting patients to resources. Many social needs are documented only in free-text clinical notes within the electronic health record, making systematic identification labor-intensive and impractical at scale. Although Large Language Models (LLMs) have shown strong performance in extracting social needs and generating clinical summaries, their capability to synthesize HRSN information from medical records has not been evaluated. There is a critical need to determine whether LLMs can accurately and efficiently summarize HRSN data to improve identification and intervention in clinical settings.

Solution

To address this gap, Duke University's Department of Medicine and the Duke Population Health Management Office (PHMO), along with the Duke Institute for Health Innovation (DIHI), formed a transdisciplinary team to develop and implement the HRSN

Discharge Summary Note Solution. At the time of a patient's discharge from the hospital, our solution automatically generates a HRSN Discharge Summary Note within Maestro Care (Duke's instance of Epic, the electronic health record) using relevant data from the patient's hospitalization and past medical and social history. The tool integrates the ChatGPT-4.1 model via Microsoft Azure, a Large Language Model (LLM), to synthesize both free-text documentation from clinical notes and structured screening data captured in flow sheets by nurse teams. It applies a standardized discharge summary template and prompt (Figure 1) to produce consistent outputs via the LLM. The evaluation of the LLM-generated output focused on hospitalized patients with diverse social needs profiles and varying admission complexity to ensure applicability across a broad range of cases. HRSN Discharge Notes were evaluated for quality on a 5-point Likert scale by physicians and clinical social workers across domains including:

- Completeness
- Accuracy
- Clarity
- Conciseness
- Credibility
- Overall usability

The findings were incorporated back into the prompt to improve the model's output across all domains.

Impact

Through the evaluation of the HRSN Discharge Summary's refined output, the solution's HRSN Discharge Summary note scored well across all domains: completeness 4.75 ± 0.49 , accuracy 4.85 ± 0.24 , clarity/readability 5.00 ± 0.00 , handling of missing data 4.90 ± 0.21 , and overall satisfaction 4.85 ± 0.24 (Table 1).



RFA PROJECT IMPACT

The HRSN Discharge Summary went live on November 1, 2025. From November 1, 2025 to March 1, 2026, HRSN Discharge Summary Notes were automatically generated for 13,491 hospitalizations across the DUHS hospitals. The Transitions Team utilizes the HRSN Discharge Summary Note in their preparation for follow up calls with patients who have identified HRSNs.

The impact of the HRSN Discharge Summary Note is improved care for more of Duke's patient population in need. Specifically, this solution increases the clinical capacity and reduces the cognitive burden for the DukeWELL Care Transitions Team. Time savings for the Population Health Care Manager and others are being collected, with current estimates of 975 hours annually.

This translates to a total cost avoidance of \$60K (hourly pay) for FTE effort being redirected to patient care.

Next Steps

The team will complete impact measurement on time savings and increased HRSN identifications/interventions for Duke patients. A clinical manuscript has been drafted and will be submitted to disseminate this work.

Academic Output

(pre-print)Prizant et al. AI-Generated Summaries of Health-Related Social Needs to Support Enhance Hospital Discharge Coordination. JAMIA 2026

Grading Domain	Mean Likert Score (SD)	Perfect Score %
Completeness	4.75 (±0.49)	70%
Accuracy	4.85 (±0.24)	70%
Clarity + Readability	5.00 (±0.00)	100%
Handling Missing Data	4.90 (±0.21)	80%
Overall Satisfaction	4.85 (±0.24)	70%

Table 1: Evaluation of Finalized HRSN Summary Notes (n=10 HRSN Discharge Summaries for 10 hospitalizations)

Health Related Social Needs Discharge Summary Note Tip Sheet

What is the HRSN Discharge Summary Note: All adult patients will now have a [Health-Related Social Needs \(HRSN\)](#) Discharge Summary Note generated in Maestro Care on the day after their discharge from the hospital. The note contains details [on](#) the patient and their recent hospitalization:

1. Patient information
2. Current Admission Information
3. Current Discharge Planning
4. Home Medical Needs Summary
5. Social Needs Summary

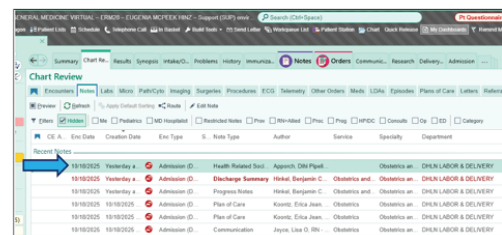
How is the note created: the note is automatically generated using AI. The AI tool scans Maestro Care for the patient's notes and flowsheet data (flowsheet data on any food, financial, housing, or transportation needs) from the previous six months. Then, it summarizes that information to create the HRSN Discharge Summary. It includes this disclaimer: "This summary was drafted using an artificial intelligence (AI) solution. It is intended to support care planning. Please verify information with the medical record or care team before making decisions."

Who is the note for: this is an "internal communication" note meant to support care managers as they prepare to follow up with [patients](#) post-discharge.

When is it available: All adult hospital discharges starting on Wednesday, 10/22/2025 will include the HRSN Discharge Summary Note

Where to find the note: In [MaestroCare](#), in the same locations as you find patient notes (e.g., Chart Review, Notes Activity):

- Note Type: Health Related Social Needs
- Note Author: [Apporch](#), [Dihi Pipeline db](#)



- Disclaimer to start each note: "This summary was drafted using an artificial intelligence (AI) solution. It is intended to support care planning. Please verify information with the medical record or care team before making decisions."

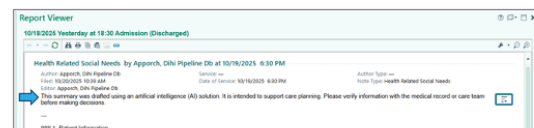


Figure 1: Training Tip Sheet

Health-Related Social Needs Discharge Summary

1. Patient Information

- **Name:**
- **MRN:**
- **Age:**
- **Gender:**
- **Primary Residence Type:**
- **County/city of residence:**
- **Caretaker/Emergency Contact Information:**
- **Power of Attorney (POA):**
- **Who the patient lives with:**
- **Insurance Information:**
- **Preferred Pharmacy:**
- **Advanced Care Planning:**

2. Current Admission Information

- **Date of Admission (DOA):**
- **Date of Discharge (DOD):**
- **Number of Days in Hospital:**
- **Reason for Admission:**
- **Hospitalizations in the Past Six Months:**

3. Current Discharge Planning

- **Discharge Location:**
- **Primary Caretaker:**
- **New Medical Needs/Referrals:**
- **New Social Needs/Referrals:**
- **Mobility Upon Discharge:**
- **New/Changed Medications at Discharge:**
- **Upcoming Appointments:**

4. Home Medical Needs Summary

- **Home Medical Supplies/Therapies:**
- **Durable Medical Equipment (DME):**
- **Home Health Services:**
- **Baseline Mobility Level:**

5. Social Needs Summary

- **Most Recent SDOH Screening:**
 - **Date:**
 - **Social need categories and results:**
 - Food insecurity:
 - Financial resource strain:
 - Housing instability:
 - Transportation:
- **Positive Findings in the Past Year:**
- **Referrals Placed for Social Resources:**
- **Resources Received by Patient:**
- **Other important patient considerations:**

Figure 2: HRSN Discharge Summary Template

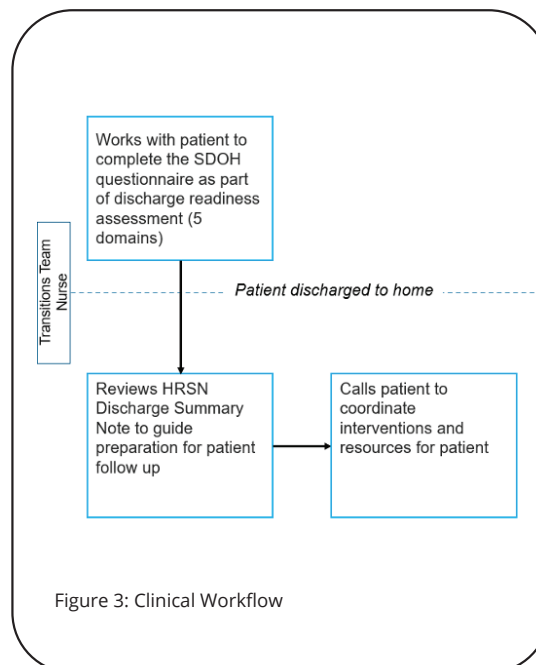


Figure 3: Clinical Workflow

Figure 2

Shows the tip sheet provided as part of the training to this team.

Figure 3

Shows the clinical workflow incorporating the HRSN Discharge Summary Note.



Automating Oncology Patient Access and Elevating the Patient Experience

Emily Norboge
Matthew Engelhard, MD, PhD
Jennifer Kennedy-Stovall, BS
Ton Wang, MD, MS
Mary Ann Plambeck, MSN, RN
Will Ratliff, MBA
Mike Revoir, BS
Suresh Balu, MS, MBA

Project Brief

Newly diagnosed oncology patients need access to the right provider quickly to establish a plan for treatment, to mitigate any disease progression and reduce stress. Duke Cancer Institute supports easy access to the right provider for these patients, via the new patient access center and a hotline number for prospective patients to call. Scheduling these patients can be time consuming due to manual chart review and complex patient cases. To support our front-line schedulers and sustain the growth of the program, we use artificial intelligence to automatically review a patient’s chart and quickly determine the appropriate provider and time window for scheduling.

Problem

Ease of access to care for newly-diagnosed oncology patients is vital for both clinical outcomes and patient experience. A majority of newly diagnosed patients experience high levels of psychological distress combined with a desire for information about their care and treatment options at the time of diagnosis.¹ Initial experience is a key

component in a patient’s choice for where they seek treatment. This is evidenced by studies that show a ‘fast track’ oncology clinic offering results in significant increases in new patient volume.² As part of new patient intake, a New Patient Coordinator (NPC) reviews the patient’s chart to obtain relevant clinical data, and then applies a set of scheduling rules to determine the correct provider and the correct timeframe for scheduling an initial appointment. This chart review is a manual, time-consuming step that limits the capacity of the patient access center to service a growing volume of prospective patients in need. Additionally, there is a substantial maintenance cost to train new NPCs to perform this scheduling task effectively. There is an opportunity to use artificial intelligence to expedite the chart review and decision-making required to determine the appropriate provider and time window for scheduling.

Solution

To address this challenge, Duke Cancer Institute (DCI) and the Duke Institute for Health

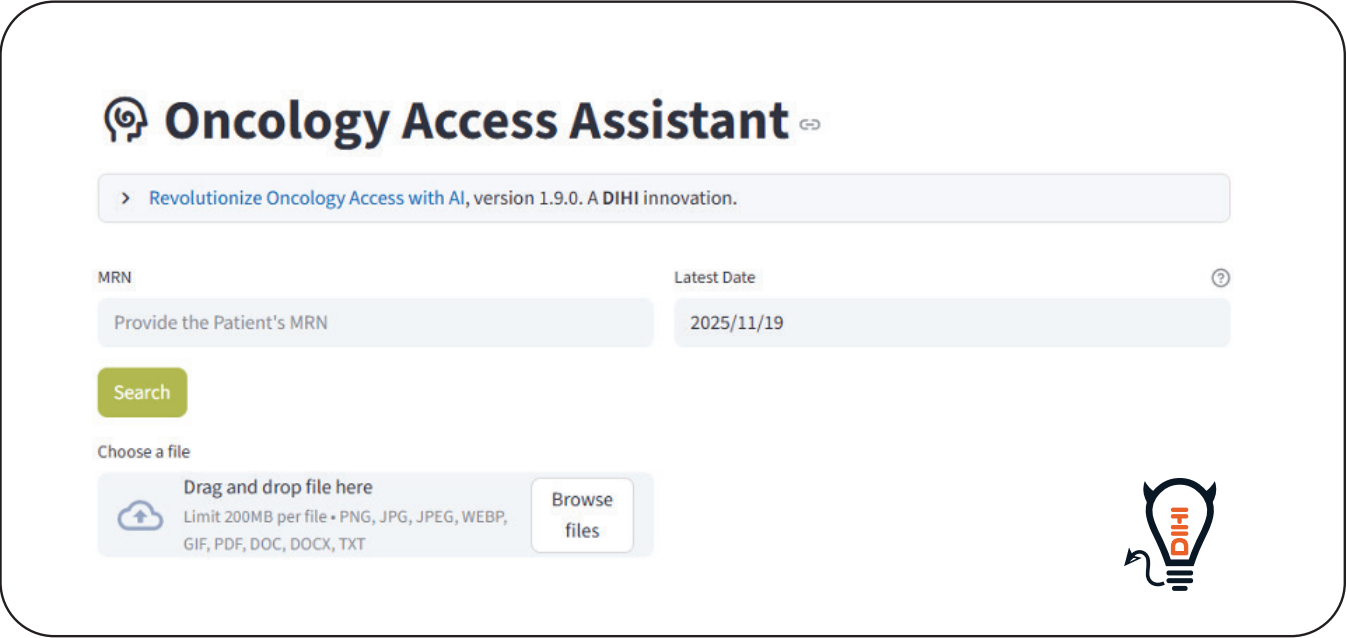


Figure 1: The Oncology Access Assistant (OAA) overview.

Innovation (DIHI) formed a transdisciplinary team to develop and implement the Oncology Access Assistant (OAA) (Figure 1). The OAA uses artificial intelligence to automatically review a patient’s chart and determine the appropriate provider and time window for scheduling. When an NPC is triaging a prospective patient, they access the OAA via a secure Web App, type in the patient’s MRN, and click “Calculate scheduling requirements”. The OAA automatically retrieves all the patient’s clinical notes, imaging result narratives, pathology results, lab results, and diagnosis codes from Maestro Care (Duke’s instance of Epic, an electronic health record) over the previous 18 months. It also searches for and retrieves notes from other health systems for the patient via the Care Everywhere functionality within Epic. The Care Everywhere data incorporation is critical for a large portion of these prospec-

tive patients, who have received some type of relevant care prior to seeking treatment at Duke. The OAA then feeds the data, along with a refined prompt that includes the scheduling rules, into a large language model (LLM), to produce instructions on which provider and within what timeframe to schedule an appointment (Figure 2). It also provides justification for those results, citing relevant clinical data used to make its determination. To evaluate the solution, 30 retrospective cases were adjudicated by clinical leads on the project team to establish ground truth “schedule with” and “schedule within” answers along with supporting data and reasoning per the scheduling logic.

NEXT PAGE →

New Breast Patient Evaluation

Schedule With: Medical Oncologist, Surgical Oncologist, Radiation Oncologist

Schedule Within: 2 weeks, with priority to Surgical Oncology

Rationale: The patient meets criteria for early stage breast cancer, ER/PR positive, HER2 negative. According to the rubric, such patients should be scheduled with Medical Oncology, Surgical Oncology, and Radiation Oncology within 2 weeks, with priority to Surgical Oncology.

Symptom or Diagnosis: Early Stage Cancer: ER/PR Positive, HER2 Negative

Relevant Documentation

Criteria: Early Stage Cancer: ER/PR Positive, HER2 Negative

Quoted Text: "Histology: invasive ductal carcinoma Grade: 2 Biomarkers: ER+, PR+, and HER2- Location: 12:00, 17 cmfn Clip shape: ribbon Size: 1.6 cm"

Source: Clinical Note, [REDACTED], Physician, 2024-08-02

Rationale: This note documents the new diagnosis and biomarker status, matching the rubric for early stage, ER/PR+, HER2- breast cancer.

Criteria: Early Stage Cancer: ER/PR Positive, HER2 Negative

Quoted Text: "Right breast 12:00, 17 cm from nipple, needle core biopsy containing invasive ductal carcinoma (SD24-9288, block A1) ER IHC 97 3+ POSITIVE PR IHC 91 3+ POSITIVE HER2/neu IHC 0 0 NEGATIVE"

Source: Pathology Note, DUHS Pathology, 2024-07-18

Rationale: Pathology confirms ER+, PR+, HER2- status, which is the key biomarker profile for this scheduling rubric category.

Start a new conversation

What would you like to ask the assistant?

Figure 2: Example output of the Oncology Access Assistant — scheduling recommendation with clinical rationale.

23

The OAA solution separately assessed the ground truth cases and produced its determination. The initial phase of the OAA solution implementation focuses on support of NPCs who triage new patients who have received a diagnosis related to breast cancer. The future state workflow (Figure 3) maintains a “human in the loop” step wherein the NPC reviews the OAA solution output before finalizing the scheduling step.

Impact

The evaluation of the Oncology Access Assistant showed a 97% accuracy of the solution. 29 of 30 cases were successfully identified by the OAA solution for “schedule with” and “schedule within” answers and the supporting rational and documentation. Assessment of the case that failed revealed ambiguous language in the documentation about the recency of imaging documentation, which was a contributing point in the scheduling logic for determining the correct answers. The OAA solution went live on March 3, 2026. At the three-week mark post-go live, the OAA solution had been used to support scheduling tasks for 77 new patients. On average, it took the OAA solution 14 seconds to provide the “schedule with” and “schedule within” answers, along with rational and supporting documentation references from the patient’s chart. This approach is more than fifty times faster than a manual review by a NPC, which takes 10-15 minutes per new

patient triage. The total time saved by using the OAA solution to review the 77 new patients’ documentation is estimated to be 15.74 hours. Using an estimated annual use volume of 3,000 cases (~8 per day), this solution achieves an annual cost savings of \$31,000 in nursing time.

Next Steps

The Oncology Access Assistant’s impact will continue to be assessed for time saved and for accuracy of scheduling recommendations. The project team is expanding the OAA to support new GI Oncology patient triage using the same methods, with a go-live goal of summer 2026.

References:

1.Whelan, Timothy J., E. Ann Mohide, Andrew R. Willan, Andrew Arnold, Michelle Tew, Scott Sellick, Amiram Gafni, and Mark N. Levine. “The Supportive Care Needs of Newly Diagnosed Cancer Patients Attending a Regional Cancer Center.” Cancer 80, no. 8 (1997): 1518–24. [https://doi.org/10.1002/\(sici\)1097-0142\(19971015\)80:83.0.co;2-7](https://doi.org/10.1002/(sici)1097-0142(19971015)80:83.0.co;2-7).

2.Basta, Y L, K M Tytgat, H H Greuter, J H Klinkenbijn, P Fockens, and J Strikwerda. “Organizing and Implementing a Multidisciplinary Fast Track Oncology Clinic.” International Journal for Quality in Health Care 29, no. 7 (2017): 966–71. <https://doi.org/10.1093/intqhc/mzx143>.

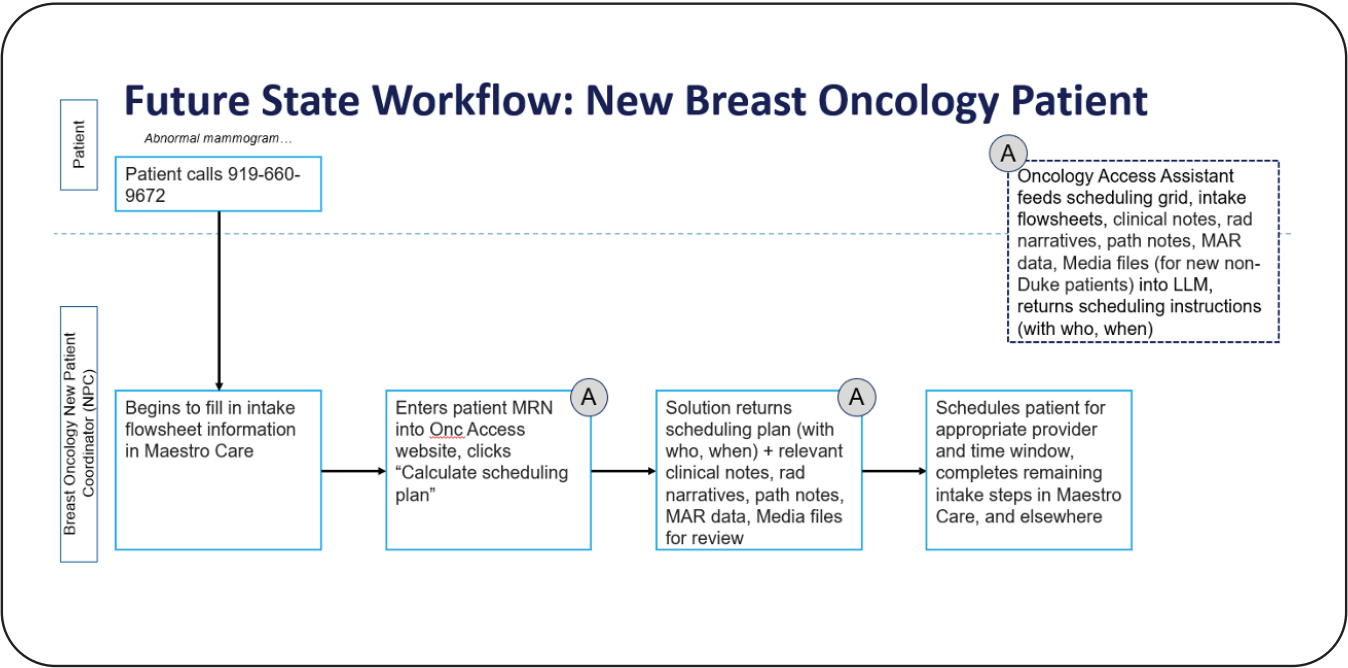


Figure 3: Future-state workflow — the OAA embedded in the New Patient Coordinator triage process.

Predicting Delayed Graft Function to Improve Kidney Transplant Matching

Stuart Knechtle, MD (PI)
 Kristen Lott, MSN
 Jacqueline Henson, MD
 Matt Ellis, MD
 Sait Tunc, PhD
 Will Ratliff, MBA
 Kartik Pejavara, BS
 Mike Revoir, BS
 Marshall Nichols, MS
 Suresh Balu, MS, MBA



Project Brief

When a deceased donor kidney becomes available for transplant, transplant surgeons must rapidly determine whether the organ is an appropriate match for a specific recipient. While every offer must go through various filters to ensure a baseline level of organ quality, the final organ acceptance decision requires synthesizing donor characteristics, recipient medical complexity, and contextual factors, often under significant time pressure. To support this decision-making process, our team is developing a machine learning model that leverages donor and recipient data to predict the risk of short- and long-term kidney dysfunction following kidney transplantation. By providing objective, data-driven risk estimates, this tool aims to improve match quality, standardize decision-making, and ultimately improve patient outcomes.

Problem

Kidney transplant offer-evaluation is complex, high-stakes, and time-sensitive. Donor information is uploaded into the United Network for Organ Sharing (UNOS) UNET system and offered to transplant centers in sequence. At Duke, the on-call surgeon must quickly assess donor kidney quality and determine whether a given recipient is likely to achieve an acceptable outcome. The current process presents several challenges, including:

- Surgeon-to-surgeon variability in acceptance practices
- Inconsistent decision-making over time
- An effort-intensive evaluation process requiring review of structured and unstructured data
- Lost opportunities to transplant usable kidneys
- Occasional use of poor-quality kidneys
- Prolonged cold ischemia time due to delayed decisions

These limitations may contribute to patient harm, inefficiencies in organ utilization, and variability in transplant outcomes. Importantly,

existing UNOS filters and structured data fields do not capture the full complexity of donor-recipient interactions, nor do they incorporate longitudinal recipient data or nuanced clinical narratives. A data-driven predictive approach offers the opportunity to support transplant surgeons with individualized risk estimates, helping optimize kidney acceptance decisions while preserving clinical judgment.

Solution

To address this problem, the Duke Transplant Center partnered with the Duke Institute for Health Innovation to develop a machine learning model that predicts short- and long-term graft dysfunction outcomes following kidney transplantation, namely 90-day graft failure, 1-year graft failure, and whether creatinine level exceeded 2.5 mg/dL 6-months post-transplant. Our goal is to provide transplant surgeons and coordinators with objective, patient-specific risk predictions at the time of organ offer, enabling more standardized and informed “go/no-go” decisions. By supporting optimal donor-recipient matching, this tool aims to improve graft outcomes, reduce unnecessary organ discard, and increase the number of successful kidney transplants performed at Duke.

We used data collected from 174,708 kidney transplant events across the United States between 2015 and 2025, curated from the United Network for Organ Sharing (UNOS) registry. The dataset was filtered for adult transplant recipients and deceased donors. Outcomes were derived from UNOS follow-up reporting. 90-day and 1-year graft failure were defined based on whether a reported graft failure date occurred within the respective time interval following the transplant date. Six-month renal dysfunction was defined using the reported recipient creatinine value at six months post-transplant, with values exceeding 2.5 mg/dL considered elevated.

Model inputs included structured donor and recipient variables capturing recipient demographics, comorbidities, and transplant status; waitlist history and dialysis characteristics; immunologic compatibility measures such as Human Leukocyte Antigens (HLA) mismatch and Calculated Panel Reactive Antibody (CPRA); donor demographics and cause of death; donor laboratory and infectious disease testing results; donor medical history and management prior to procurement; and transplant logistics including cold ischemia time and organ preservation methods. The model was trained using light gradient boosting machine (lgbm) with 70% of the data used for training and 15% each for validation and testing. Model performance was evaluated using sensitivity, specificity, and the area under the receiver operating characteristic curve (AUC-ROC).

Impact

Of the 174,708 kidney transplant events included in the modeling cohort, 3,723 (2.13%) were followed by 90-day graft failure, and 5,409 (3.10%) were followed by 1-year graft failure. For the analysis of renal dysfunction at six months post-transplant, the cohort was restricted to 152,892 kidney transplant events due to incomplete follow-up laboratory measurements in a subset of recipients. Within this subset, 7,262 (4.75%) recipients had a reported creatinine level greater than 2.5 mg/dL six months following transplantation. Model performance varied across the three predicted outcomes. Because graft failure and severe renal dysfunction are relatively low-prevalence events, the area under the precision-recall curve (AUPRC) provides a more informative measure of model performance than traditional discrimination metrics alone. AUPRC reflects how well the model identifies true positive cases relative to the expected baseline performance defined by outcome prevalence.

The model predicting 90-day graft failure achieved an AUPRC of 0.0454, representing approximately 2.1 times the baseline prevalence of the outcome. The 1-year graft failure model achieved an AUPRC of 0.059, approximately 1.9 times the baseline prevalence. The model predicting elevated creatinine (>2.5 mg/dL) six months post-transplant achieved an AUPRC of 0.130, approximately 2.7 times the baseline prevalence of the outcome. These results demonstrate that the models can identify transplant cases with elevated risk of graft dysfunction at rates substantially higher than random classification. Together, these findings suggest that donor-recipient characteristics captured in the UNOS registry contain meaningful predictive signal for clinically important post-transplant outcomes. The models provide a foundation for future clinical decision support tools that could assist transplant teams in evaluating donor-recipient matches and identifying higher-risk transplant scenarios (Table 1).

Next Steps

We are currently working with collaborators at the Duke Transplant Center to identify and prepare the appropriate data infrastructure required to retrieve donor and recipient information from the United Network for Organ Sharing (UNOS) in real-time. This will enable the predictive models to be executed at the time of organ offer, providing transplant surgeons and coordinators with risk estimates that can inform clinical decision-making during the donor evaluation process, as outlined in Figure 1. In parallel, we are continuing to refine and expand the model feature set to evaluate whether we can construct more clinically meaningful variables to further improve predictive performance. Ultimately, our goal is to integrate these models into the transplant evaluation process to support more consistent donor-recipient matching decisions and improve kidney transplant outcomes.

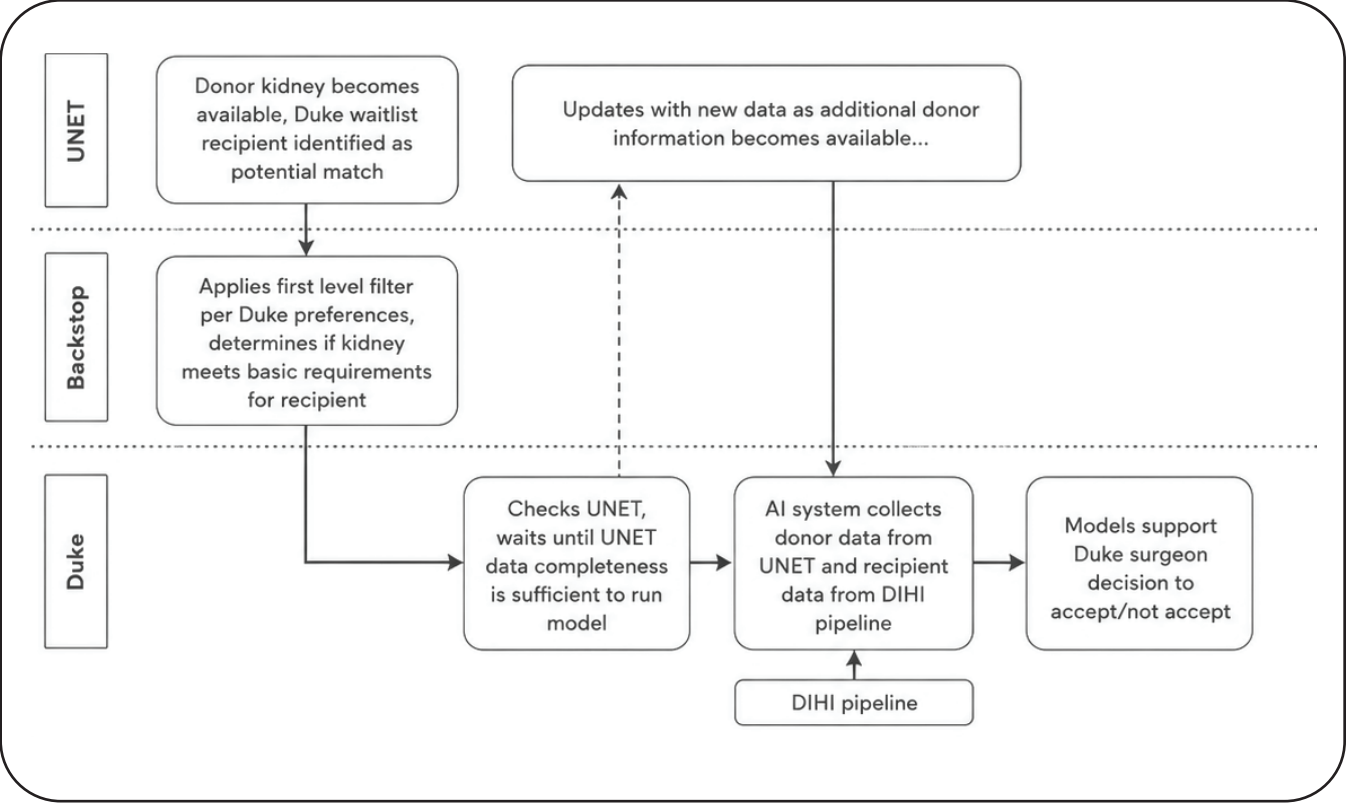


Figure 1: Kidney transplant matching prediction model — future-state clinical decision support workflow.

Model	Description	AUC	AUPRC
90-day graft failure	Risk of graft failure within 90 days post-transplant	0.6705	0.045
1-year graft failure	Risk of graft failure within 1 year post-transplant	0.659	0.059
Creatinine >2.5 mg/dL at 6 mo.	Risk of elevated creatinine indicating dysfunction	0.751	0.130

Table 1: Predictive model performance across three kidney transplant outcomes.



Spotlights

Scout: Your Thinking Partner

Will Knechtle, MBA, MPH
Michael Gao, PhD
Suresh Balu, MBA, MS



Problem

Health Delivery Organizations need information fluidity and local innovation.

Healthcare professionals face a growing crisis of information overload. More than three million articles are published annually in biomedical journals, and the data required to assess a single patient has become enormous. The Duke Institute for Health Innovation (DIHI) found that compiling a new patient's information before an appointment can take up to two hours. Meanwhile, virtualization has multiplied patient messages and administrative meetings, and many providers still rely on faxes, scanned documents, and paper records alongside their digital systems.

Time and resources are both scarce. As patient volume grows faster than the healthcare workforce, providers must access and share extensive data quickly and reliably. Data fluidity is not a convenience, but a necessity. In high-stakes clinical situations where every minute counts, comprehensive chart review could become a casualty of time pressure, forcing providers to make decisions with incomplete information. Academic medical centers are simultaneously under intense pressure to reduce costs and

optimize processes, with margins that leave little room for inefficiency. Research and support cuts have further reduced the resources institutions typically use to optimize workflows and address these challenges.

The Core Issue

Electronic Health Records (EHRs) are now ubiquitous in both hospital and ambulatory settings. Their size has become its own problem. "Information overload," "chart bloat," and "note bloat" are terms now common in the clinical literature, driven by redundancy, copy-paste habits, automated templates, and regulatory requirements.

This is not merely an administrative inconvenience but a systemic issue impacting patient safety, physician well-being, and the patient-physician relationship. Simulation studies demonstrate that high cognitive load from conventional, data-heavy EHR interfaces is associated with four times as many errors compared to streamlined displays that show only salient information. When critical data are buried in a cluttered chart, a physician's situational awareness suffers.



The bloat of the typical EHR has restructured medical work itself. To avoid disrupting the patient relationship during clinic hours, many physicians defer documentation evenings and weekends, a phenomenon often called “pajama time”. In our study, 400 care providers reported an average of 4.23 hours of pajama time per week (median: 3, SD: 4.21) and an average of 10.43 hours of EHR time within the workday (median: 10, SD: 6.0). Thirteen percent reported spending ten or more hours per week of their personal time with EHRs. High EHR burden is a primary predictor of physician burnout, decreased career satisfaction, and intent to leave practice.

Solution: Scout

Scout is an AI-powered platform for searching and synthesizing patient EHR data. Scout reviews every note, imaging report, pathology report, and other structured data in the patient’s record to deliver exactly the information a care practitioner needs, in the format they want. Scout helps care practitioners save time, reduce cognitive burden, and maintain or exceed care quality standards.

Scout functions like an AI assistant, similar in interaction style to ChatGPT, but instead of searching the web, it searches across a patient’s medical record. It has been designed with patient safety and data provenance as foundational principles. Every response Scout provides is traceable back to a specific location in the patient’s record, and our system architecture is built to minimize the risk of inaccurate or unsupported output.

We are studying whether Scout outperforms traditional AI tools with features such as

- Full record review with effective EHR integration
- Efficient data processing to enhance speed and minimize unnecessary tokens
- Easy-to-use and granular data provenance
- User-controlled queries
- Simple and loveable design

Early internal testing across thousands of queries has shown:

- A near-zero rate of unsupported or fabricated responses
- High accuracy across a range of query types and clinical domains
- Consistent data provenance, allowing clinicians to verify Scout’s responses directly

Scout will create value across the full care spectrum, supporting clinical, operational, administrative, financial, research, and academic work.

Impact

- Complete tasks that normally take hours in just 1-3 minutes, reducing clinical burden and improving efficiency.
- Significant reduction in mental demand, effort, and temporal demand as measured by the NASA Task Load Index.
- A primary conclusion of a study (the prospective, randomized, evaluator-blinded, two-period crossover trial publication is forthcoming) was that Scout maintained quality, meeting “non-inferiority” criteria in accuracy, completeness, and relevance compared to manual EHR workflows.

LLM Operations: A Framework for Standardizing Clinical Chart Pipelines in Healthcare

Marshall Nichols, MS

Introduction

Clinical notes have long been considered one of the most information-rich but least standardized data sources in healthcare. They are now being read, reorganized, and rewritten by AI-driven systems. Managing these systems responsibly requires a new set of practices. Think of it like quality control on an assembly line: someone must make sure every step produces reliable, safe output. In healthcare AI, that discipline is called LLM operations, or LLM Ops. (LLM = Large Language Model).

The challenge is that clinical data originates from many disparate sources: electronic health record (EHR) systems, laboratory systems, radiology platforms, and external health information exchanges. Each system may store and represent data differently, even when describing the same clinical concept. Without a standardized foundation, LLM-based pipelines risk amplifying inconsistency, introducing bias, or producing outputs that cannot be reliably used by other systems.

What healthcare AI needs is a common language. It needs one standard all systems can read and write, regardless of where the data originated. Fast Healthcare Interoperability Resources (FHIR) has emerged as that. FHIR is a shared rulebook that tells healthcare software how to package and label information so other systems can reliably understand it. By providing a common data model, terminology bindings, and application programming interfaces (APIs), FHIR offers a pathway for normalizing clinical inputs before LLMs process them, and for expressing LLM outputs in a form that downstream systems can consistently interpret.

The Role of LLM Ops in Clinical Note Pipelines

LLM Ops provides the operational framework for managing the full lifecycle of AI-driven clinical documentation.

It encompasses model selection, prompt design, versioning, performance monitoring, error handling, and compliance auditing. In healthcare, these concerns are heightened by the need for explainability, traceability, strict access controls, and security.

A standardized data layer is essential for reducing these risks. When LLM pipelines operate on loosely structured or proprietary data formats, every integration point becomes a potential error source. Normalizing inputs and outputs through a common standard like FHIR allows LLM Ops teams to define clear contracts between systems, simplifying validation, testing, and governance. Training or prompting against predictable data structures is more reliable than working around an evolving patchwork of ad hoc formats.

Benefits of Using FHIR for Standardization

Semantic Interoperability

FHIR organizes clinical data into defined building blocks called resources. Each resource covers a specific type of information: a patient's identity, a lab result, a diagnosis, a medication order, or a clinical note. Because every system uses the same building blocks, an AI model can work with data from a hospital in Ohio the same way it works with data from a clinic in California. When clinical notes are derived from these resources, their meaning becomes portable across systems.

For LLM pipelines, this means structured context can be provided alongside free text in a predictable format. An LLM summarizing a patient encounter can be supplied with standardized data that clarifies vital signs, diagnoses, and lab results. The resulting note can then reference those same standardized elements, making it easier for other applications to parse, validate, or build on the output.



Modular and Scalable Architectures

Each stage of the process, collecting data, cleaning it, transforming it, and sending it where it needs to go, works with the same standardized format. Think of shipping containers: regardless of what is inside, every container is the same size and shape, so every port, crane, and truck can handle it without custom modifications. A pipeline might ingest raw data from multiple EHR vendors, map it into a unified set of FHIR resources, and store it in a centralized repository. An LLM service can then generate standardized clinical notes from those resources, published back into the ecosystem for analytics platforms, care coordination tools, or patient portals to use.

Improved Governance and Auditability

Healthcare AI solutions must demonstrate how clinical data is processed and transformed. FHIR supports this through provenance records, essentially a paper trail showing where data came from, who touched it, and how it changed. When LLM outputs are wrapped in FHIR resources, those provenance records can indicate which model version, prompt configuration, and input data produced a given note. This traceability is critical for clinical safety reviews, regulatory audits, and internal quality assurance.

Challenges in Integrating FHIR with LLM Pipelines

Data Quality and Completeness

FHIR defines a common structure but does not guarantee complete or consistent data. Different organizations may populate only a subset of fields, use local codes instead of standardized terminologies, or represent the same clinical concept differently. Models may perform well with richly populated resources but degrade when key fields are missing. Building robust pre-processing layers is essential, but adds complexity and operational overhead.

Mapping Unstructured Text to Structured Resources

Much clinical information remains in free-text notes. Converting it into structured FHIR resources requires natural language processing or manual abstraction, both of which can introduce errors. LLMs are frequently used for this extraction, which creates a compounding risk: if the AI makes a small error pulling information from a note, that error gets baked into the next set of data the AI learns from. Over time, small mistakes can quietly grow into larger ones. Continuous monitoring, sampling, and human review are essential to catch these issues before they affect clinical workflows.

Performance and Latency

Consider a physician seeing a patient right now. If the AI tool meant to assist that visit takes thirty seconds to load because it is processing too many data checks, it becomes a burden rather than a help. Speed is not just a technical preference in health-care. It is a patient safety issue. Caching strategies and asynchronous processing can help, but they add layers that must themselves be managed and monitored.

Regulatory and Privacy Considerations

Standardizing clinical notes through FHIR does not eliminate obligations related to patient privacy, consent, and data sharing. In some cases, standardization makes it easier to distribute information more widely, which increases risk if access controls are not properly enforced. FHIR provides mechanisms such as security labels and consent resources, but implementing them consistently across LLM pipelines and downstream applications remains a significant undertaking.

Designing a FHIR-Centric LLM Ops Framework

A practical approach begins with defining clear data contracts at each stage of the workflow. Raw inputs from source systems should be mapped into a canonical set of FHIR resources representing the minimum required clinical context, forming the basis for both LLM prompts and validation rules.

Monitoring and governance should be built around FHIR's metadata and provenance features. Performance metrics, error rates, and quality indicators can be linked to specific resource versions and model deployments, creating a feedback loop for continuous improvement.

Conclusion

Standardizing clinical note pipelines through LLM Ops and FHIR offers a clear pathway for improving interoperability, governance, and scalability in healthcare AI. By grounding AI-driven transformations in a common data model, organizations can reduce integration complexity, improve traceability, and make AI-generated insights accessible across a broad ecosystem of applications.

The convergence of LLM operations and FHIR is not a distant possibility. It is already taking shape in health systems building governed, standardized pipelines today. The organizations moving deliberately on this now will not just keep pace with healthcare AI. They will help define what responsible deployment looks like.

What We've Built and How to Keep It

Will Knechtle, MBA, MPH

Indisputable impact

Scout reduces cognitive load for care practitioners by 40% and cuts EHR-task time by 40%. Sepsis Watch reduced mortality by 25% at Duke University Hospital. eConsult reduced provider burnout by 24%. Duke Health Policy Chat, born as a DIHI innovation, has become integral to Duke's institutional policy platform. These outcomes are real, measurable, and increasingly visible across Duke Health.

Less visible is where they came from.

They came from a system: an integrated, compounding, deeply embedded system that took a decade to build and that grows more valuable with every project it produces. This article is about understanding that system and what can quietly damage it.

The system, not the credentials

Duke people are genuinely accomplished. The DIHI team combines advanced degrees, deep clinical knowledge, and technical expertise that would be competitive anywhere in the country. But what makes DIHI work is not credentials alone. It is credentials combined with relational skills, carefully cultivated personality mixes, and healthcare knowledge that runs deep enough to understand not only what a workflow is supposed to be, but how workflows genuinely operate at two in the morning on a short-staffed floor.

Researchers call this profile "pi-shaped": professionals with deep expertise in two domains who also excel at cross-functional collaboration. DIHI hires deliberately for this combination. The flat hierarchy that results creates an environment where a technical question gets answered in under an hour, where solution architects challenge clinicians in the same weekly session where data scientists review each other's work, and where no single person's recognition matters more than the collective result. Research across industries consistently shows

that teams built around relational depth outperform teams built around individual stars, and that organizations relying on high performers without collaborative culture see innovation decline over time. DIHI's results confirm this at scale.

The physical infrastructure compounds this advantage. The Monorail data pipeline processes EHR data every five minutes, running millions of real-time algorithms invisibly beneath clinical workflows. This represents a decade of engineering that cannot be purchased off the shelf or reconstructed quickly. Now, DIHI is building its successor: EHRx, a FHIR-native data infrastructure designed for the interoperable future of healthcare. Like Monorail before it, EHRx is not just an internal tool. It is protectable intellectual property with independent commercial value, reflecting DIHI's growing capacity to build foundational infrastructure rather than isolated applications.

New hires require approximately eighteen months to become fully productive. They must learn DIHI's coding structures, build clinical relationships across service lines, and integrate into workflows refined over years. The system itself is the asset. Individual brilliance, without the integrated system surrounding it, delivers a fraction of what the system makes possible.

DIHI's clinical relationships deepen this further. Across every service line at Duke, DIHI has developed a hyper-local understanding of how care is delivered in each unit: who the informal decision-makers are, where the workflow friction lives, and how change gets adopted or rejected. These relationships took years to build. They enable the rapid iteration and genuine clinical adoption that distinguish DIHI's tools from tools that succeed in pilots but die in deployment.



This is also how DIHI structures its work. Roughly ten projects are chosen annually through a competitive process that seeks alignment between executive priorities and frontline pain points (where the C-suite's concern and the night-shift nurse's work-around point to the same underlying problem). Each partnership runs for twelve structured months: four months building the team and mapping workflows, two months preparing data, three months developing and testing in semi-simulated environments, and the final months preparing for real deployment and measuring impact. Every voice carries equal weight regardless of credentials. The data scientist and the attending physician are partners—a mechanism by which solutions get used in the health-care world.

What can damage it

DIHI is among the rare 20% of teams that research identifies as sustaining high-functioning collaboration over time. This is a result of intentional culture-building, consistent infrastructure investment, and hiring decisions that consistently prioritized collaborative fit alongside technical excellence. Each of those three dimensions can be undermined through well-intentioned decisions that seem locally reasonable but are structurally disruptive.

Introducing hierarchy or differentiated rewards shifts motivation subtly but decisively from collective performance to individual recognition. Once that shift begins, the relational safety that makes DIHI's sessions work (where anyone can challenge anyone, where questions get answered in under an hour) starts to erode.

Hiring for credentials over collaborative fit introduces people who are individually accomplished but who pull against the culture rather than into it. This is the hardest mistake to see coming and one of the hardest to reverse. The eighteen-month onboarding reality means a poor cultural fit does not reveal itself immediately. By the time it does, the cost in team cohesion is already paid. The corollary is equally important: the people best positioned to evaluate collaborative fit are the ones who have built and lived inside the culture. Hiring decisions made at a distance, by people unfamiliar with DIHI's specific operating model, carry a structural risk that good intentions cannot offset.

Structural reorganization (absorbing DIHI into a larger department, changing reporting lines, or splitting the team across separate institutional priorities) severs the embedded clinical relationships that make rapid iteration possible. DIHI's effectiveness depends on being a coherent, neutral, trusted partner to every service line simultaneously.

Finally, vendor relationships without clear IP frameworks create a specific and underappreciated risk. Care practitioners, accustomed to teaching and publishing, naturally share knowledge openly. In the context of sophisticated enterprise technology partnerships, it can mean that workflow intelligence built over years is absorbed into external products without equivalent exchange. Hence, the question every Duke Health leader should ask before authorizing a deep vendor engagement is not whether the partner is reputable, but whether the exchange is genuinely mutual: what are they providing that we lack, and what are they requesting that represents institutional knowledge we built?

NEXT PAGE →



When the answer to both questions is clear and balanced, partnerships create real value. DIHI's licensing relationship with Cohere Med illustrates what this looks like in practice. Cohere Med brought global business infrastructure and international deployment capacity (things DIHI genuinely lacks and cannot quickly build). DIHI brought healthcare-specific expertise, a user-tested and fully functioning technology, and frameworks for responsible AI deployment. Neither party could replicate what the other contributed. The result was something that neither could have built alone, with clear agreements governing IP ownership, data governance, and ongoing responsibilities.

What Duke has, and what sustaining it requires

Scout, Sepsis Watch™, Bed View, eConsult, Duke Health Policy Chat, and now EHRx did not emerge from individual brilliance or from any single vendor relationship. Rather, they emerged from an integrated system combining proprietary data infrastructure, hyper-local clinical knowledge, a collaborative culture built over a decade, and the institutional trust to deploy solutions that clinicians use in practice.

This is a strategic asset that exists few other places in this configuration. It creates competitive barriers that cannot be quickly replicated and organizational learning that compounds with every project. Sustaining it requires two things from Duke Health leadership, and both matter equally.

The first is active protection: awareness of the decisions that quietly erode the system, and the discipline to ask, before reorganizing a team or signing a partnership agreement or making a hire, whether that decision reinforces the integrated model or fragments it.

The second is productive support: resourcing DIHI in ways that trust the team's own judgment about what it needs. The people who built this system are the people best equipped to identify the talent, infrastructure, and management structures that will carry it forward. Support that provides resources while deferring to that expertise compounds DIHI's value. Support that substitutes external judgment for internal knowledge, however well-meaning, risks dismantling what it intends to strengthen.

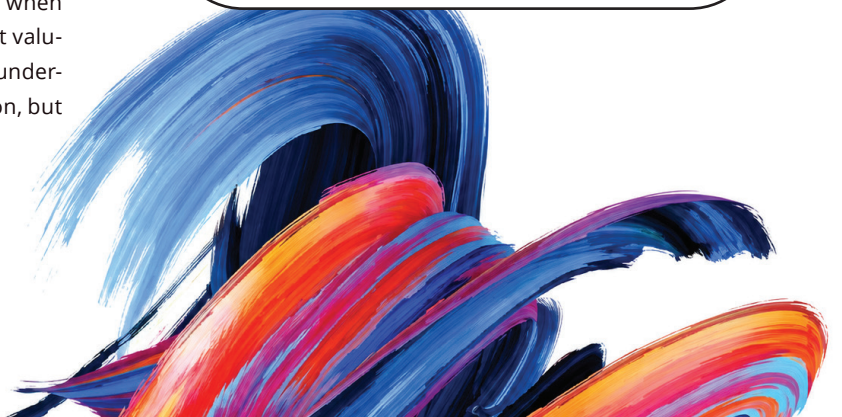
DIHI's first decade demonstrates what becomes possible when an organization treats its collaborative culture as its most valuable asset. The second decade depends on leadership understanding that the same culture requires not just protection, but the kind of support that trusts the team to know itself.

How DIHI Delivers Value at the Point of Care

For a clinician on a high-acuity floor, value is not found in a boardroom slide deck; it is found in a safer, more manageable shift. Consider, for example, the transition of care in the CTICU. At the end of every shift, attending physicians previously navigated a systemic hazard: the manual summary of critically ill patients on mechanical circulatory support or awaiting transplant, completed under significant cognitive fatigue, often led to inconsistent detail and varied urgency.

When vital data is buried in a wall of text, situational awareness erodes and patient safety may be compromised. DIHI addresses this burden directly by deploying AI and real-time infrastructure to surface the right information at the precise clinical moment it is needed.

What distinguishes DIHI's impact is a full-stack philosophy that prioritizes sustainability over isolated successes. When a solution's logic proves effective in the CTICU, it is engineered to be portable (to the MICU, to primary care, to any unit facing structurally similar information hurdles). Every deployment strengthens a shared digital foundation that grows more capable with each iteration. It is the cumulative effect of these factors, rather than any single tool, that enables a dedicated team to generate results far beyond its size.



Measuring the ROI of Healthcare Innovation



Shared responsibility

Every member of our team owns a piece of this. When a new tool reaches the floor, its value depends on clinical leads, engineers, and program managers ensuring it gets used, measured, and improved. These frameworks exist so anyone can think clearly about what their project is worth and how to prove it.

TWO COMPLEMENTARY METHODS



TOP-DOWN

System-level attribution

Start with institution-wide metrics (cost, utilization, mortality) and attribute improvement back to deployed innovations.

$$\Delta \text{ outcome} \times \text{volume} \times \text{unit value} = \text{estimated system impact}$$



BOTTOM-UP

Project-level measurement

Measure each solution at the point of care (time saved, errors avoided, patients served) then roll up to a portfolio total.

$$\text{unit savings} \times \text{encounter count} = \text{project-level ROI}$$



CORE ROI FORMULA

$$\text{ROI} = ((\text{Benefits} - \text{Costs}) \div \text{Costs}) \times 100\%$$

$$\text{Benefit} = \Delta \text{ metric} \times \text{volume} \times \text{unit value} \times \text{confidence}$$

SIX ROI LEVERS IN HEALTHCARE INNOVATION



Clinician time recovery

$$\Delta \text{ min/encounter} \times \text{encounters/yr} \times \text{hourly rate}$$

Scout reduces cognitive load and EHR task time. eConsult linked to reduced provider burnout.



Adverse event avoidance

$$(\text{events avoided} \times \text{penalty avoided}) \pm \text{DRG revenue delta}$$

Sepsis Watch reduced mortality at DUH. Penalties and quality losses may outweigh DRG gains.



Revenue capture

$$\text{denied claims recovered} \times \text{avoided denials} \times \text{avg. claim value}$$

Prior Auth AI achieved near-perfect accuracy; 340B tool projects significant annual savings.



Access expansion

$$\text{new patient encounters} \times \text{margin/encounter}$$

Onc Access Assistant reduced time-to-appointment for high-risk oncology referrals.



IP & licensing value

$$\text{licensing revenue} + \text{build-vs.-buy delta} + \text{vendor risk avoided}$$

EHRx: FHIR-native pipeline with independent commercial potential. Cohere Med license.



Education & research spillover

$$\text{publications} \times \text{citation value} + \text{trainees} \times \text{career premium}$$

Over 10 years of RFA projects. DIH work contributes to a significant share of Duke's peer-reviewed output.

FOUR VALUE PATHWAYS: FOLLOW THE CHAIN



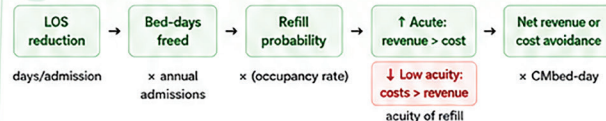
Clinician time recovery



Saved time only generates value if redirected. Not all of it does — that adjustment makes the estimate credible.



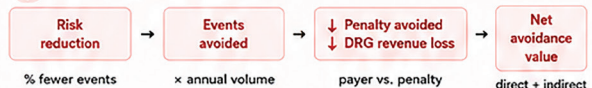
Length-of-stay reduction



A freed bed still carries fixed costs. Acuity of the incoming patient determines whether refill generates margin.



Adverse event avoidance



A complication can raise DRG payment — but CMS penalties and quality-driven volume losses may outweigh it.



Revenue capture and protection



Accuracy at the documentation step is what makes everything downstream possible.

TWO OFTEN-OVERLOOKED MEASUREMENT TOOLS



Statistical process control

Healthcare improvement rarely happens in isolation. Overlapping initiatives, seasonal variation, and regression to the mean all compete for credit. SPC charts distinguish genuine signal from noise, but even they cannot fully resolve which intervention caused which outcome. For that, causal inference methods (including counterfactual frameworks developed by Judea Pearl) offer a more rigorous path toward attributing impact to a specific project.

From correlation toward causation



Publications & citations

Peer-reviewed publications are a proxy for knowledge durability and field influence — a form of ROI that funders and faculty promote as and committees recognize. Equally important: involving medical students and trainees in AI research attracts talent to Duke, signals institutional leadership, and generates tuition revenue, prestige, and long-term faculty pipeline capacity that compounds well beyond any single project.

Knowledge that compounds over time



Jay Swayambunathan, MD, Class of 2026

DIHI Clinical Research and Innovation Scholar

The DIHI Clinical Research and Innovation Scholarship gave me the privilege of working at the intersection of clinical practice, data science, and health systems innovation — one of the most formative experiences of my medical education. In 2025, as a DIHI Clinical Research and Innovation Scholar, I brought three years of prior experience as a data scientist and a Master of Engineering in Artificial Intelligence to a range of meaningful projects. DIHI provided the structure and access that made it possible to apply that background to real clinical problems, connecting me with the teams, tools, and environments where that work could have impact.

My DIHI projects spanned diverse areas, each teaching me important lessons about the challenges and opportunities of bringing new technologies into healthcare. My primary project involved leveraging large language models to automate the review of prior authorization documentation. DIHI's positioning at the intersection of clinical operations and technical research meant this project was grounded in genuine institutional need rather than academic abstraction. This work deepened my understanding not only of the technical capabilities of these models, but also of the administrative and economic dimensions of medicine. Ensuring accuracy across highly variable documentation was a major challenge, and we developed innovative methods to address it.

We also validated an ensemble model designed to predict adult inpatient deterioration. It was DIHI's collaborative environment (bringing data scientists and clinicians into the same room) that made it possible to pair rigorous technical validation with real clinical context, ensuring that models are not only accurate but also deliver real-time value at the bedside.

Another project had me assessing an educational dashboard aimed at enhancing feedback and evaluation for medical students. This highlighted how thoughtful design and data-driven tools can enrich the experience for learners and educators alike. As a medical student, it was especially rewarding that DIHI created space for me to contribute to the culture of innovation within the Duke University School of Medicine.

Beyond these individual projects, my time at DIHI reinforced the importance of interdisciplinary collaboration. By design, DIHI brings together clinicians, engineers, and administrators around shared problems. I witnessed firsthand how clinicians, engineers, and administrators bring unique but complementary perspectives to produce solutions no one could reach alone. This lesson will remain central to my future career. I learned that innovation in healthcare is ultimately not just about technology. It is about people, systems, and the pursuit of better care for all patients. I am deeply grateful to the DIHI team for their mentorship and support, and I look forward to carrying these lessons into the next chapter of my clinical journey.





Srijan Bhasin, MD, Class of 2026

DIHI Clinical Research and Innovation Scholar

My experience at the Duke Institute for Health Innovation (DIHI) began shortly after returning from my research year in Singapore; and it quickly became one of the most formative chapters of my medical training. DIHI offered the ideal environment to expand upon my longstanding interest in computational methods and their applications in healthcare. It was the first place where I was able to take a project from conceptualization to full workflow implementation within a large health system, giving me unmatched insight into the innovation process within healthcare.

My primary project focused on improving hand-off communication in the cardiothoracic ICU using large language models. The goal was to create a predictable, structured, and efficient way for providers to exchange critical information during shift changes in a high-acuity environment. Achieving this required true interprofessional collaboration. Working alongside clinical researchers, project managers, solutions architects, and healthcare administrators forced me to think deeply about each component of clinical operations. Multidisciplinary thinking and engagement skills will remain essential as I continue within healthcare as a trainee.

The DIHI experience pushed me to grow in new directions that built upon my background in statistics and traditional machine learning. DIHI guidance on obtaining buy-in from end users and stakeholders – an essential step for sustainable impact – was arguably more important than my becoming familiar with advanced AI tools. The DIHI journal club especially reinforced that no one in healthcare works in isolation; every successful innovation requires trust, shared vision, and collaboration across disciplines.

By the end of the year, I felt confident identifying problems independently and designing solutions. So much so that my co-scholars and I initiated a project to create a medical education tool outside of DIHI's traditional project structure. This experience, paired with the community found within the institute, shaped how I think about leadership, teamwork, and responsible innovation in medicine. DIHI not only broadened my technical and clinical perspectives, they also taught me how to ask better questions, critically evaluate new technologies, and navigate the dynamics of implementing change in healthcare settings.

I leave DIHI with a stronger sense of purpose and a network of mentors and colleagues who I know will remain integral to my career in medicine.





Nikkie Dutta, MS

DIHI Intern, Master's Program, Dept of Population Health Sciences

I joined the Duke Institute for Health Innovation (DIHI) as an intern during the second year of my master's program in Fall 2025. The experience has been one of the most rewarding of my academic career. While I've held several positions in the pharmaceutical industry outside of academia, this was my first time being involved in an end-to-end project within a large hospital system. As someone passionate about improving care delivery and access, DIHI gave me a new perspective on how innovation can be operationalized at scale.

The project I've been working on is titled "Utilizing AI to Reduce the Burden of Prior Authorizations." The goal is to automate the review of patient medical records to generate prior authorization requests. We initially focused on four high-volume medications: Ocrevus, Botox, Pluvicto, and Darzalex Faspro. Once a patient with one of these medication requests is identified, the tool collects a patient's prior medical history, extracts relevant data from clinical notes, and packages that information into a prior authorization request.

One of my biggest takeaways has been understanding the nuances of training a model to reliably produce expected outputs. For example, I validated ground truth files for Botox to ensure the model was trained on accurate, representative data, while simultaneously refining prompts to improve the experience for clinicians.

This taught me that building a trustworthy AI tool isn't just about the model itself but requires understanding the entire workflow, from scanning and extracting clinical data to generating outputs that stakeholders can confidently use.

Participating in stakeholder meetings deepened this understanding further. Seeing the downstream impact on denial teams, payers, providers, and patients reinforced the importance of iterative, user-centered design. Presenting our model outputs in real time and problem-solving alongside end-users surfaced challenges I never would have identified working on the tool in isolation.

Beyond the technical work, I've genuinely valued DIHI's mentorship culture and the team's shared commitment to meaningful innovation. As an intern, I was immediately welcomed and supported in a collaborative environment that fostered learning and growth. I found it very valuable that team members actively offered their insights and guidance to help me further develop my skills as a student researcher. As I approach Spring 2026 graduation, I'm looking forward to finishing out the semester with the team and am excited to leverage these experiences in my future work.





Tareq Aljurf, MD

Pulmonary & Critical Care Fellow, Duke University

I joined the Duke Institute for Health Innovation (DIHI) in early 2025, and my experience with them has been phenomenal. At the time, I was a third year pulmonary and critical care fellow. While I had always enjoyed working on personal projects involving software and programming, I had never been part of a healthcare project with a real technology focus. I was looking for real life hands-on experience at that intersection, and that is what led me to DIHI.

Right off the bat, the team was warm and welcoming. After exchanging introductions and learning about my background in critical care, they recommended I join the ICU Handoff project, which was already underway. The project's goal was to use large language models to automatically summarize clinical notes into concise handoff summaries for clinicians, condensing a process that typically takes over 30 minutes. The team onboarded me quickly, helped me get up to speed on their progress, and we dove right in, experimenting with different approaches to generate the best possible outputs.

What I found especially valuable during my time on the project was the openness of the broader DIHI team. I regularly reached out to members beyond our immediate project group to learn

from their expertise and work through challenges as they came up. The project also gave me a unique opportunity to apply my clinical knowledge and exposure alongside modern large language models. DIHI taught me ways to interact with variable inputs and missing data that is simply the reality of hospital environments.

From there, I became involved in other projects across DIHI. Sometimes I contributed to major components; other times, my role was more consultative, such as lending clinical context that team members from non-medical backgrounds might not have. What many clinicians don't realize about DIHI is that it's fundamentally built on collaboration. Your clinical knowledge genuinely matters here. But just as importantly, you find outstanding colleagues and resources, many who are experts in fields you may have never encountered before.

The range of projects is broad, and there is room for most interested clinicians at all levels of technical experience. Ultimately, that openness is what makes DIHI special: it gives clinicians the opportunity to grow and contribute well beyond the typical boundaries of their training and practice.





Mark Sendak, MD, MPP

Co-Founder & CEO, Vega Health, Former DIHI member

I spent the last 11 years working full-time at the Duke Institute for Health Innovation (DIHI), before launching Vega Health as the company's CEO. My decade at DIHI included the opportunity to work on over 100 projects and mentor nearly 100 students and trainees. In short, DIHI transformed my career – and I wouldn't be where I am today without what I learned here. And it all started with Suresh Balu asking if I might take a year off from medical school to join the team.

I owe so much to DIHI, that I thought I'd pass along a few of the most important lessons from my time here:

Lessons for Students

Lesson 1: Higher education tries to force you into a tradeoff between specialization and broad expertise. They're wrong.

As an undergraduate student at UCLA and graduate student at Duke, I had extremely broad interests spanning mathematics, economics, physics, computer science, statistics, public policy, business, and medicine. Many faculty and advisors recommended I transition from exploring interests across domains to exploiting interests within specific domains.

I resisted. Between 2015 and 2025 I contributed to the development and validation of over 60 AI solutions. The scientific discipline of AI and machine learning in healthcare is now established. Ten years ago, it wasn't. What was perceived as undisciplined exploration as a student fifteen years ago became deep specialization in a now well-established field.

Lesson 2: But...there is a tradeoff between the prestige of expertise and growth opportunities.

Many of my peers in medicine sought opportunities to work with prestigious faculty in prestigious fields at prestigious institutions. These opportunities were hyper-competitive and those who secured these positions quickly established themselves as experts.

Working at DIHI, we rarely (if ever) functioned as domain experts. Every year, each project required us to enter new business units across Duke Health as outsiders and we had to quickly understand complex workflows, organization structures, and business models. We deeply respected the contextual nuances within the system and elevated the perspectives of front-line workers closest to the problems – because we had to.

Each DIHI project served as an opportunity to grow and learn. By the time we built expertise within a particular domain, we were completing projects and soliciting proposals for the next innovation cycle. The pace of innovation kept me and my colleagues humble to the complexities of the problems we worked on and the value different Duke colleagues brought to the table.

Lesson 3: Give yourself time to learn, build skills, and cultivate mentors

At the Machine Learning in Health Care conference in 2015, a well-known investor told me I should leave medical school to start a company. While I didn't seriously consider that option, I did understand the allure of leaving school to build a company.



It took me another 10 years to spin Vega Health out of Duke. Over that time, our team at DIHI deliberately built capabilities and demonstrated value across a broad portfolio of AI technologies. We built the Health AI Partnership national collaborative to surface and disseminate best practices for AI implementation and provide technical assistance to community hospital systems. The scar tissue from internal and external implementation experiences accumulated.

Deliberately investing in my growth between 2015 and 2025 allowed me to then move very fast. Over the last eight months, Vega Health has raised \$5M in seed financing, recruited a stellar team, closed two health system contracts and completed one implementation. The skills, credibility, and network cultivated over the last ten years set me up for success.

Lessons for Healthcare Innovators

Lesson 1: Integrate into, don't deploy against

The language we used at DIHI to describe our work had significant impact on the way we were perceived within the organization. While we were entrepreneurial technologists, we were also colleagues to the front-line clinicians caring for sick, complex patients. If technology implementations failed, we would be held accountable. DIHI's work would cease if we were not seen as a valued partners to front-line workers.

We spent significant time and energy understanding the organizational and contextual nuances of the settings in which we worked. For each project, we would shadow front-line clinicians to directly observe workflows and we would interview a broad range of stakeholders. We had to understand the root cause of problems and the organizational context of the implementation setting before any technology was introduced.

For this reason, we focused on integrating technologies into the organization rather than deploying technologies against the organization. This distinction was reflected back to us by an anthropologist who previously studied technology deployments in the military. Technology is often framed as being deployed to disrupt, a stark contrast to our more measured and nuanced approach of integrating technologies into Duke Health.

Lesson 2: Don't just select talent; grow it

DIHI's work cut across many domains that require specialized skills and training. The technologies we built were complex. We often leveraged emerging machine learning methods on high-dimensional, large datasets. The meaning and interpretation of

individual data inputs often required specialized clinical domain knowledge. Team members building the models were different than the team members who understood the data.

The workflows and operations of business units in which the technologies were implemented were equally, if not more, complex. These factors were best understood by front-line workers with direct experience in the implementation context. These team members were also different from those who understood the models or the data.

At DIHI, we assigned trainees and students to work in these intersections. These trainees joined interdisciplinary teams to develop a deep understand of the workflow, data sources, and modeling methods. Equipped with this rare experience, many DIHI alumni were highly sought after for industry and academic positions.

Lesson 3: Never get too far from the problem

The best technologies are built by those who most deeply understand the user pain point and who can most quickly iterate on the solution. In healthcare, this requires working in a context that is highly aligned with the front-line worker. Front-line workers must trust you before they'll ever expose deeper layers of the problem. They must trust you to keep giving you feedback on versions of the solution.

For healthcare innovators, this often means working within provider organizations. While provider organizations may not have the budgets and as many engineering talents as software companies, they serve as fertile ground for product design and development. When the solution is built and validated within the provider organization, then scale the technology with an external commercial partner.

What's Next

Vega Health will continue to complement the work of DIHI and Health AI Partnership. Through a commercial vehicle, we will scale innovations built by the DIHI team (and other innovators) to community and rural settings most affected by the digital divide. We will continue to provide students and trainees opportunities to engage in our work and we aim to scale the impact of healthcare innovators everywhere.

AI Governance & Ethics

Health AI Partnership (HAIP): Program Overview

Jee Young Kim, PhD
Alifia Hasan, B. Pharm, MBA
Sena Kpodzro, MPH
Ciera Thomas, MPH
Suresh Balu, MS, MBA

About HAIP

Launched in 2021, the Health AI Partnership (HAIP) is a community of practice comprising more than 35 healthcare delivery organizations and ecosystem partners. HAIP brings together leaders across the health sector to share insights, develop best practices, and advance the safe, ethical, and effective use of artificial intelligence (AI) in healthcare.

HAIP's mission is to empower healthcare professionals to use AI effectively, safely, and ethically through community-informed, up-to-date best practices. Its vision is to be the trusted partner and up-to-date source of actionable guidance for healthcare professionals using AI. HAIP is guided by its core values: (1) eliminating the digital divide, (2) improving patient care, (3) improving the healthcare workplace, and (4) building community.

Five Pillars of the HAIP Program

1. Engage with the Community of Practice

HAIP's foundation is its vibrant community of healthcare organizations, researchers, and partners committed to advancing responsible AI adoption. This year, HAIP expanded its Corps Network and launched its first Practice Network cohort, engaging four Federally Qualified Health Centers (FQHCs) and one Community Health Center (CHC) across the U.S. in a 12-month AI adoption program that concluded in September 2025. The program supported participants in exploring and implementing AI solutions aligned with their operational and clinical priorities.

To inform future scale-up, HAIP also interviewed about 40 safety-net and health IT network (e.g., Health Center Controlled Networks) leaders in California. The interviews provided systematic insights into the state's AI landscape, including high-value use cases, regional challenges and needs, and AI readiness.





2. Share Knowledge and Experience

Through public and members-only events, HAIP connects innovators, implementers, and policymakers to exchange insights and lessons learned from real-world AI adoption. The Health AI Hub and Real-World AI Series continue to serve as key venues for this exchange. In 2025, HAIP partnered with the Center for Care Innovations (CCI) to launch a new webinar series, “AI in Action.” The series spotlights research and implementation of AI-based tools addressing the unique needs of FQHCs and safety-net healthcare organizations, with the goal of advancing equitable access and improved patient outcomes.

3. Advocate for the Community of Practice

HAIP also serves as a collective voice for healthcare professionals and safety-net organizations, advocating for regulatory clarity, evidence-based standards, and investment in responsible AI adoption. By engaging policymakers and aligning with national initiatives, HAIP ensures that the perspectives of community-based providers shape emerging AI governance frameworks.

4. Co-develop Health AI Resources

HAIP collaborates with its partners to create practical, actionable tools that help organizations evaluate and implement AI responsibly. In the past year, HAIP collaborated with Roche to co-develop a two-stage framework for evaluating healthcare AI products, enabling organizations to assess technical performance, clinical relevance, and commercialization readiness. Together with the American Medical Association (AMA), HAIP also designed an AI ethics training program for physicians, equipping physicians to apply ethical principles in everyday AI use. In addition, HAIP released

a vendor disclosure framework to support healthcare leaders in making transparent, informed procurement decisions. These tools are designed to bridge the gap between high-level principles and operational decision-making in AI adoption.

5. Disseminate Knowledge and Learnings

To amplify community insights, HAIP actively contributes to the broader conversation on AI in healthcare through publications, media, and thought leadership. A case study documenting AI governance development within a large Canadian hospital system was published this year, while the new evaluation framework and AI ethics training are under review for publication. HAIP members also share learnings through conference presentations, media commentary, and opinion columns, extending the partnership’s impact across national and international audiences.

Looking Ahead

Building on these achievements, HAIP will continue to grow its network, expand the Practice Network program to new regions, and strengthen its role as a trusted source of practical, responsible AI guidance for healthcare professionals.

Through its collaborative, community-driven approach, HAIP remains committed to ensuring that every healthcare organization—regardless of size, location, or resources—can safely and effectively harness AI to improve care, eliminate digital divide, and build a stronger health system.



The Health AI Partnership Practice Network: Mitigating the Digital Divide in AI Adoption

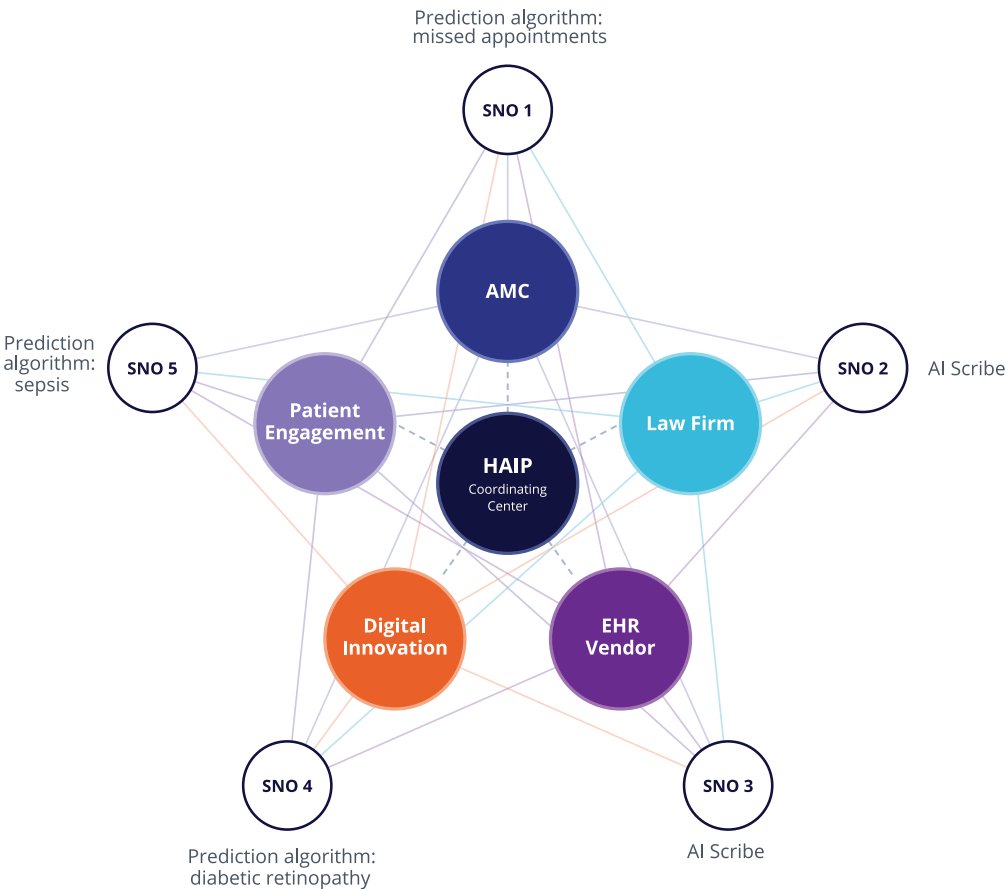
Ciera Thomas, MPH

Introduction

The adoption of AI tools in healthcare has been the focus of a growing body of research due to its potential to address operational inefficiencies, improve the safety and quality of patient care, and alleviate physician burnout. Existing research suggests that Safety Net Organizations (SNOs) may stand to benefit significantly from the use of AI tools, yet they exhibit lower rates of AI adoption compared to other health delivery organizations (HDOs). This trend reflects the growing digital divide, or the gap between individuals, households, businesses, and geographic areas at different socio-economic levels regarding their opportunities to access information and communication technologies. With SNOs serving over 31 million patients annually, understanding and addressing this gap is essential to prevent the exacerbation of existing health disparities for the underserved patients that SNOs predominantly care for.

Health AI Partnership (HAIP) sought to address this gap by building on lessons learned from prior technical assistance (TA) programs, such as Regional Extension Centers to support EHR adoption. We launched a 12-month TA program—the Practice Network—to support SNOs through AI adoption and lifecycle management. Through this approach, we aimed to:

- 1. Guide SNOs through AI adoption and lifecycle management,
- 2. Understand high-impact AI use cases for SNOs, and
- 3. Identify and address challenges associated with AI implementation in SNO settings.



The Program Model

The Practice Network employed a distributed hub-and-spoke model in which SNOs function as "Spokes," receiving technical assistance for AI adoption, while organizations with AI adoption experience serve as "Hubs," providing guidance. HAIP served as the Coordinating Center, acting as both a Hub providing technical assistance and the organizing body responsible for identifying participants, facilitating connections, and managing day-to-day program operations. Ten additional Hubs were drawn from HAIP's Community of Practice and brought specialized expertise across clinical, technical, operational, regulatory, and ethical domains—including HDOs with prior AI implementation experience, an international law firm, an EHR vendor, a healthcare innovation nonprofit, and a community-based organization. Five Spokes—four FQHCs and one community hospital—were SNOs across the U.S., each pursuing first-time adoption of a specific AI tool. These tools spanned multiple use cases, including ambient AI-powered scribes, a predictive model for no-show appointment risk, a sepsis prediction model, and an AI-assisted screening tool for diabetic retinopathy.

The Practice Network was launched in September 2024 and ran through September 2025. Using the HAIP Key Decision Point framework, the Coordinating Center developed customized implementation plans for each participating Spoke, focused

on four critical phases of the AI lifecycle: Procurement, Development/Adaptation, Integration, and Lifecycle Management. Technical assistance was tailored to participants' needs and delivered via one-on-one mentorship, virtual office hours, and peer learning forums.

Barriers and Strategies for AI Adoption

Although all five SNOs made measurable progress over the course of the Practice Network, they encountered significant challenges spanning technical, operational, and financial domains. These challenges highlight the systemic barriers faced by resource-constrained organizations pursuing first-time AI adoption.

1. Limited foundational AI knowledge

A lack of familiarity with AI concepts and tools among both staff and leadership was a widespread issue across participating Spokes. This knowledge gap delayed adoption timelines and created uncertainty about how to use AI effectively within clinical workflows. As part of the Practice Network, Spokes implemented tailored team training initiatives. These trainings, designed in collaboration with Hubs, provided practical overviews of AI functionality and its role in clinical and operational settings. Supplemental peer learning opportunities allowed participants to exchange insights across organizations, helping accelerate progress.

NEXT PAGE →



2. Challenges in evaluating AI tool performance and impact

Spokes reported difficulty systematically assessing how AI tools performed in real-world conditions, both during pilot testing and after rollout. Lacking clear frameworks for evaluation, organizations struggled to interpret vendor-provided data and determine whether tools met their unique needs. Hubs provided targeted mentorship on developing metrics to measure success, including tool accuracy, predictive utility, and impact on clinical efficiency. Through guidance on both baseline data collection and ongoing monitoring, Spokes were able to refine their evaluation processes and strengthen decision-making.

3. Inadequate strategies for communicating AI use to patients

Communicating the introduction and purpose of AI tools to patients posed a significant challenge, especially in settings serving economically and socially vulnerable populations where trust in healthcare services can already be fragile. To address this, all Spokes ultimately chose to disclose AI use to patients. With input from Hubs, Spokes received guidance on considerations like data privacy, trust-building, and regulatory compliance when developing disclosure processes. Several Spokes leveraged focus groups and surveys to better understand patient concerns, enabling the design of effective communication strategies that reinforced transparency and accountability.

4. Financial constraints

The cost of adopting and sustaining AI tools, including upfront licensing fees, infrastructure upgrades, and long-term tool maintenance, was a major barrier across all Spokes. These financial constraints were compounded by limited negotiation power when engaging with AI vendors. Hubs helped Spokes build ROI analyses tailored to their operational and patient care goals, equipping organizations with data to make the case for securing funding. While some Spokes pursued external grants to supplement adoption costs, financial barriers persisted, highlighting a need for broader systemic solutions.

5. Lack of AI governance processes

Most Spokes lacked internal governance systems to navigate ethical and compliance concerns related to AI adoption. Questions around algorithmic bias, data privacy, and long-term oversight of AI systems created hesitation when adopting tools. Spokes worked with Hubs to establish governance processes uniquely suited to their organizational structures. This included the creation of cross-functional teams tasked with overseeing AI implementation and ensuring adherence to regulatory standards and ethical principles. These efforts proved critical in ensuring organizational accountability.

Impact and Sustainability

The Practice Network demonstrated that safety net organizations (SNOs) can make meaningful progress in adopting AI tools when supported by structured mentorship and TA. By centralizing specialized knowledge and distributing it through tailored guidance, the program provided resources that would have otherwise been inaccessible to participants, addressing critical gaps in workforce readiness, governance, and evaluation processes. At program completion, three of five Spokes had successfully piloted their AI tools, with the remaining two preparing for pilots by the end of 2025. Post-program surveys revealed a 53% increase in organizational AI knowledge, with 89% of Spoke teams rating their knowledge as medium or high.

Despite this progress, the current model's reach is limited. Thousands of SNOs across the U.S. face similar barriers and scaling this model is critical to meeting widespread needs. To this end, HAIP is working to expand the Practice Network with future cohorts designed for regional and use-case-specific engagement. In California, HAIP has engaged more than 30 stakeholders, including health center networks, hospital associations, and managed care plans, to better understand capacity and barriers to adoption. These efforts lay the groundwork for scaling the network to meet the needs of diverse care settings nationwide.

Expanding the Practice Network offers a roadmap for bridging the digital divide in healthcare and ensuring AI technologies can be effectively integrated into SNOs. Sustained investments in infrastructure, capacity building, and program design are essential to extend the impact of AI in the safety-net ecosystem. Through continued partnerships, state-specific initiatives, and grant-funded programs, HAIP is committed to enabling SNOs to implement AI tools that enhance clinical care and improve operational outcomes.



Building Trust Through Transparency: The HAIP AI Vendor Disclosure Framework

Sena Kpodzro, MPH

The Challenge

As artificial intelligence (AI) systems rapidly diffuse into health services, healthcare delivery organizations (HDOs) face a critical gap: unclear AI-system evaluation criteria and insufficient vendor disclosures. Without a structured best-practice framework, procurement teams struggle to assess AI systems comprehensively. This leads to accountability gaps, integration challenges, and potential patient safety risks. The complexity extends beyond algorithm performance to encompass data stewardship, technical integration, lifecycle management, and regulatory compliance. These are all areas where vendors and HDOs often lack shared vocabulary. Without clear guidance, HDOs may not adequately communicate their informational needs, and vendors may be uncertain about what is expected of them.

Solution

We recruited four volunteers from three U.S.-based HDOs within the Health AI Partnership (HAIP), a national learning collaborative that brings together multidisciplinary experts from HDOs and ecosystem partners to co-design a new community-informed AI vendor disclosure framework. The framework provides HDO vendor engagement teams with structured guidance across five essential assessment domains. The HAIP AI Vendor Disclosure Framework is designed for vendor engagement teams at HDOs and AI vendors, supporting procurement, evaluation, and implementation by surfacing critical information for safe, effective, and ethical adoption of AI systems.

The five framework domains:

1. System Capabilities and Intended Use (foundational functionalities and stakeholders),
2. System Performance and Compliance (operational metrics, biases, and regulatory status),
3. Data Stewardship (governance, security, and retention policies),
4. Integration Requirements (technical prerequisites and total cost of ownership), and
5. Lifecycle Management (ongoing vendor support and maintenance responsibilities)

Together, these domains ensure HDOs have a comprehensive picture of any AI system before procurement. Each assessment domain includes:

- a domain description,
- a checklist of minimum requirements to establish baseline transparency, and
- descriptive statements that guide vendors in shaping their disclosures.

These statements tell vendors exactly what information is expected, prompting comprehensive, context-specific responses. This structure ensures that disclosures are both consistent and actionable, enabling healthcare delivery organizations to make well-informed decisions about AI procurement and oversight.

NEXT PAGE →



Impact

The framework establishes clear responsibilities between AI vendors and HDO teams, preventing accountability gaps and building trust through transparent communication. By expanding disclosure beyond algorithms to include infrastructure, safety features, and implementation expertise, it enables vendor engagement teams to make evidence-based procurement decisions aligned with organizational governance and patient safety priorities. Clear documentation of system capabilities, limitations, performance metrics, data security practices, and adverse event protocols helps reduce misuse, misunderstanding, and potential harm. It also supports proactive identification of risks across diverse patient subgroups and implementation contexts. Designed to be scalable, the framework can be integrated into formal AI governance structures at large academic medical centers or adapted through flexible working groups in smaller or resource-limited HDOs.

By fostering collaborative networks and shared resources, it helps build capacity for responsible AI adoption while reducing the risk of deepening the digital divide.

Table 1 maps these real-world use cases to the HAIP Key Decision Points (KDPs) - a structured approach that identifies eight decision points to consider when implementing an AI system.

Next Steps

This work complements HAIP's Model Facts Labels for frontline clinicians and supports compliance with the HTI-1 Final Rule mandating algorithm information disclosure. Ongoing feedback from end users will be critical in maintaining its clarity and rigor while improving its practicality.

To learn more about the vendor disclosure framework, please visit healthaipartnership.org

Framework Use Case	Relevant HAIP Key Decision Points
Guide initial vendor screening	KDP 2.2: Evaluate internal resources · KDP 2.3: Assess buy-vs-build · KDP 2.4: Assess quality of external AI products
Support the RFP process	KDP 2.4: Assess quality · KDP 2.6: Audit investment · KDP 3.1: Define performance targets
Establish baseline performance expectations	KDP 3.1: Define performance targets · KDP 3.2: Define successful use · KDP 5.1: Locally validate prior to integration
Inform contract & SLA negotiations	KDP 2.5: Assess legal risks · KDP 3.1–3.2: Performance targets · KDP 5.2: Identify and mitigate risks
Guide data sharing and access protocols	KDP 2.5: Assess legal risks · KDP 4.1: Design workflow · KDP 6.3: Prevent inappropriate use
Establish internal performance monitoring	KDP 3.1: Define targets · KDP 7.1–7.2: Monitor performance & work environment · KDP 7.5: Sustain outcomes
Define testing and validation procedures	KDP 5.1: Local validation · KDP 5.2: Mitigate risks · KDP 7.3: Audit AI solutions
Support regular performance reviews	KDP 7.1: Monitor performance · KDP 7.4: Identify risks · KDP 8.1–8.2: Determine update or decommission

Table 1 • Mapping HAIP Key Decision Points (KDPs) to vendor disclosure framework implementation use cases.

Development of Solution Facts Labels for Healthcare Generative AI

Tyng Yeh, MD, PhD

Background and Motivation

Healthcare institutions are rapidly adopting generative artificial intelligence (GAI) technologies that fundamentally differ from traditional predictive models. While predictive AI systems output bounded probabilities or classifications, generative AI creates novel content, from clinical documentation to administrative text and patient communications (He et al., 2025; Rouzrokh et al., 2025). However, these systems' capacity for open-ended outputs introduces unique challenges, including hallucinations and potential disclosure of protected health information (Kim et al., 2025; Chen & Esmailzadeh, 2024).

Colleagues at DIHI previously developed the Model Facts Label to standardize transparency for predictive AI models in healthcare (Sendak et al., 2020). This framework, inspired by pharmaceutical labeling, effectively communicated model characteristics to clinical users. However, GAI's stochastic outputs, multi-component architectures, and diverse applications necessitate substantial framework expansion.

Framework Development

Our interdisciplinary team developed the Solution Facts Label by systematically integrating requirements from academic guidelines, regulatory frameworks, and practical implementation needs. The framework consists of modular components that document the complete GAI solution ecosystem. The Mechanism section classifies systems by application domain (clinical, operational, or administrative) and specific task type, providing essential context for risk assessment. The Performance Assessment module allows flexible specification of evaluation methodologies—whether expert clinical review, automated benchmarking, or hybrid approaches—recognizing that different applications require distinct validation strategies. Additional sections address Risk and Warnings, and Governance, ensuring transparent communication about sys-

tem limitations, mitigation strategies, and monitoring procedures.

Multi-Stakeholder Transparency

A key feature is the Solution Facts Label tiered visibility structure that calibrates information to stakeholder needs. Each framework element is tagged for four visibility levels: (1) Public Registry for comprehensive regulatory disclosure; (2) Institutional Administration for governance and oversight requirements; (3) End-User Clinical Staff for practical operational guidance; and (4) Patient for plain-language summaries. This stratified approach ensures regulatory bodies receive complete technical documentation, clinicians obtain actionable operational information, and patients understand AI involvement in their care without requiring technical expertise.

Current Status and Future Directions

The Solution Facts Label represents ongoing research in healthcare AI transparency, with a comprehensive manuscript currently in preparation. This forthcoming publication will detail the complete methodology, provide expanded implementation guidance, and present additional validation across diverse GAI applications. The framework addresses the critical gap between academic reporting standards, regulatory requirements, and practical implementation needs in healthcare settings.

As generative AI transforms healthcare delivery, transparency frameworks must evolve correspondingly. The Solution Facts Label provides a structured yet flexible foundation for this evolution, supporting responsible GAI deployment while maintaining accountability to all stakeholders—from patients to providers to policymakers. By operationalizing transparency throughout the AI lifecycle, we aim to enable healthcare organizations to harness generative AI's benefits while preserving the trust and safety fundamental to healthcare delivery.



References

1.

Chen, Y. & Esmaeilzadeh, P. (2024). Generative AI in Medical Practice: In-Depth Exploration of Privacy and Security Challenges. J Med Internet Res, 26, e53008.

2.

He, R., Cao, J. & Tan, T. (2025). Generative artificial intelligence: a historical perspective. Natl Sci Rev, 12, nwaf050.

3.

Kim, Y. et al. (2025). Medical Hallucination in Foundation Models and Their Impact on Healthcare. medRxiv, 2025.02.28.25323115.

4.

Rouzrokh, P. et al. (2025). A Current Review of Generative AI in Medicine. Curr Rev Musculoskelet Med, 18, 246–266.

5.

Sendak, M. P., Gao, M., Brajer, N. & Balu, S. (2020). Presenting machine learning model information to clinical end users with model facts labels. npj Digit Med, 3, 41.

Basic Information

Item	Description	Information Provider	Registry	Admin	End User	Patients
Solution name	Official solution name as shown in the EHR catalogue or regulatory filings.	Developer	V	V	V	V
Developer name and contact	Name of the legal entity that built the solution, with email or URL for support.	Developer	V	V	V	V
Regulatory approval status and dates	List regulatory approval bodies and pathways (E.g., FDA De Novo) or note "Research-Use-Only". State the date approved or last review.	Developer	V	V	V	
Version number and update cycle	Current version identifier and schedule of future updates.	Developer and Implementer	V	V		
Summary of functionality	One-sentence, plain-language statement of what the AI produces and why it helps. Avoid jargon.	Developer	V	V	V	V

Mechanism

Item	Description	Information Provider	Registry	Admin	End User	Patients
Application and task domain	Choose application and task domains from Table 2.	Developer	V	V	V	V
Intended use	Who will use the tool, on what data, and for what decision or workflow step. Must match regulatory labeling.	Developer	V	V	V	V
Contraindication	Situations or populations where the model should not be used.	Developer	V	V	V	V
Intended user types	Roles authorized to use the model.	Developer	V	V	V	



Expanding The Portfolio

Scaling AI in Healthcare Begins with People, Not Code

Anusha Prakash, MSc

In August, the MIT GenAI Divide article grabbed headlines with a sobering statistic: 95% of generative AI projects in business fail, despite \$30–40 billion in investment. The failures, it argued, stem not from “bad AI” but from brittle workflows, poor integration, and limited memory. Healthcare has long known this truth. For years, studies have shown that many AI solutions in electronic health records (EHRs) stumble not because the models are flawed, but because they’re dropped into environments without fully understanding workflows, integration challenges, or the validation needed for wide adoption.

Despite decades of progress in interoperability standards like LOINC, ICD-10, and SNOMED CT, the true variability in healthcare isn’t in the codes, it’s in the context. An emergency department doesn’t function like an outpatient clinic: clinicians document differently, patient populations bring unique comorbidities, socioeconomic realities, and cultural nuances. Furthermore, every institution operates under its own practice norms and resource constraints. This living, breathing complexity is what makes scaling AI in healthcare so

challenging. The real test isn’t whether a model is accurate; it’s whether it can adapt, integrate, and thrive within the distinct environment of each health system.

Scaling AI and machine learning in healthcare isn’t just a technical challenge; it’s a data reality check. When models are built on data shaped by the unique ecosystems of each health system, their outcomes inevitably mirror those local variabilities and patient distributions. At the Duke Institute for Health Innovation (DIHI), we tackle this head-on through our Machine Learning Data Quality Assurance (ML-DQA) framework – a continuous validation process built on three pillars: completeness, plausibility, and conformance. These checks ensure that what feeds our models truly reflects clinical reality, not noise. In healthcare, “garbage in, garbage out” isn’t just a cliché, it’s a costly truth. By embedding ML-DQA into every stage of model development, validation, and monitoring, we safeguard the trustworthiness and impact of AI systems that clinicians can rely on, not just admire.



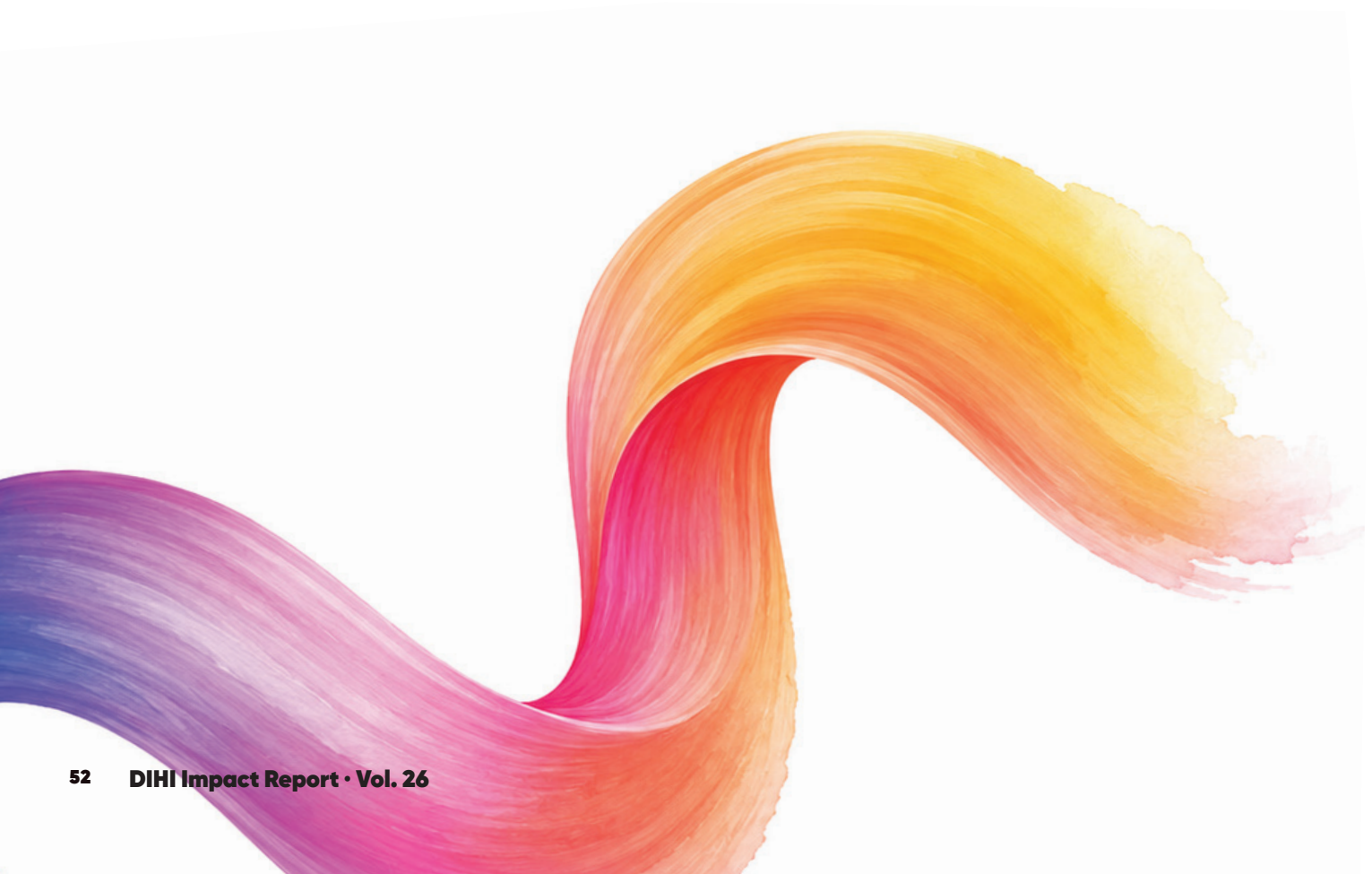
NEXT PAGE →

EXPANDING THE PORTFOLIO

Sepsis Watch™, a deep learning algorithm developed at Duke by the DIHI team and now licensed to Cohere Med, was recently validated externally at Summa Health in Ohio. This was a significant test of its generalizability across a new patient population and health system. The results, published in *Nature Digital Medicine*, demonstrated impressive robustness. But behind those promising numbers lies the real story: the months of collaboration, data preparation, and workflow alignment it took to make the validation possible. In truth, the AI itself only accounts for about 15% of what makes a deployment succeed. The remaining 85% requires understanding people, data, and processes. In healthcare, the breakthrough isn't just in the model, it's in the ecosystem that enables it to work.

In parallel, our team partnered with Epic and the University of Wisconsin to evaluate an open-source framework for local model validation: the AI Trust and Assurance Suite. Using Epic's readmission prediction model, which has been deployed at both Duke and Wisconsin since 2017, we demonstrated how this tool enables health delivery organizations, both resource-rich and resource-limited, to independently validate machine learning models within their own environments. The study showed that such open-source approaches can make model validation scalable, transparent, and reproducible across institutions, and the findings were presented at the Machine Learning for Healthcare (MLHC) Conference earlier this year.

For the past two years, our team at DIHI has been leading one of the most ambitious data efforts yet, developing a predictive model across two major health systems in Louisiana as part of an NIH-funded collaboration to expand PrEP access and reduce new HIV infections in the Deep South. This project exemplifies what true multi-stakeholder work looks like, bringing together clinicians, data scientists, community-based organizations, and health departments to bridge data with real-world impact. Over eight months of intensive work, the team has been wrangling complex EHR data, resolving entities (or 'groupers'), and engineering features that can generalize across diverse patient populations while grappling with significant data gaps in social and economic drivers of health. Beyond building a model, this stitches together fragmented data and human insight to create a solution that is equitable and scalable. Scaling AI in healthcare isn't about chasing smarter algorithms; it's about building smarter systems around them. That means blending data, empathy, and design to make AI not just accurate, but worth adopting.



Inbasket Messaging with Large Language Models

Kartik Pejavara, BS
Michael Gao, PhD
Suresh Balu, MS, MBA

Problem

Electronic health record (EHR) adoption has improved access and communication, but it has also created a growing burden of patient portal messages. Within Duke Endocrinology, hundreds of MyChart messages arrive each week, many administrative in nature but routed directly to physicians or advanced practice providers. To address this, the division established a nurse-led triage team to intercept all incoming messages, respond when appropriate, and forward others to the right destination. While this model protects providers from unnecessary inbox volume, it has shifted a substantial workload onto the triage nurses. High message volumes and expanding triage responsibilities have led to backlogs and delayed responses. These are especially problematic when they are responses to patient concerns. Nurses must review every message before determining where it belongs, even when it concerns scheduling, prior authorizations, forms, or other non-clinical requests. This manual sorting process consumes valuable time and limits the team's ability to focus on urgent patient needs.

Solution

Our team partnered with the Endocrine Specialty Engagement Center to develop an AI-assisted classification system for inbasket message routing. Working closely with triage nurses, we defined message categories reflecting real clinical workflows and built a flexible framework that predicts the most appropriate destination for each incoming patient message. The system uses large language models to interpret message intent and assign it to the correct operational category. A key strength of this approach is flexibility. Categories can be easily refined or expanded as clinical workflows evolve, enabling continuous improvement as operational and patient care needs change.

Spring 2026

Over the past year, we partnered with Duke Health Technology Solutions and endocrine triage leadership to integrate the model directly into Epic. The system now runs in real

time within the Endocrine Specialty Engagement Center InBasket pool, automatically labeling each message with a predicted category. Triage nurses can see this classification before opening the message, giving them immediate context about its likely purpose.

While automated routing remains in development, real-time labeling has already changed how the team manages messages. Messages can be sorted by predicted category, allowing staff to batch similar requests and quickly surface time-sensitive concerns. The tool operates quietly in the background at a very low cost of \$8-\$10/month.

Impact

Since implementation, the triage team reports meaningful improvements in efficiency and responsiveness. Nurses can rapidly identify high-priority messages, flag cases requiring return calls, route administrative requests without delay, and preview message intent before review. Now, nurses are spending their attention where it matters most rather than on manual sorting. The team has also seen a significant reduction in message turnaround time. Average response time has dropped from roughly three days to one. Faster responses improve patient experience, accelerate access to medications and services, and reduce the operational strain that builds after weekends or high-volume periods.

Next Steps

We are now working with the triage team to formally evaluate real-world performance and quantify time and effort savings. Ongoing nurse feedback is helping refine categories, address edge cases, and ensure high accuracy across operationally important message types. In parallel, we are collaborating with Epic to enable automated routing for low-complexity messages such as scheduling, prior authorizations, and forms. We are also exploring opportunities to bring AI-assisted triage to additional specialties, with solutions tailored to their unique clinical workflows.



Scaling Adult Deterioration Detection

Will Ratliff, MBA
Suresh Balu, MS, MBA

The Duke Institute for Health Innovation (DIHI) is at the forefront of care support for critically ill patients in Duke's hospitals. The Adult Deterioration prediction solution went live in June 2025 and is currently scaling across the Duke Health System. This go live followed a successful prospective validation of the solution output over the calendar year 2024, which showed strong performance (Table 1, Table 2).

For each hospitalized patient on an intermediate/stepdown unit, the solution applies a machine learning model to real-time data from Maestro Care, generating a risk prediction of the patient's clinical deterioration within the subsequent 12 hours. Every 20 minutes, the solution pulls new data, refreshes the risk prediction, and sends an automated page notification if the patient's risk of deterioration is above a pre-determined threshold. All notifications for that patient are then snoozed for the subsequent six hours to mitigate alarm fatigue and allow for new data to be collected to inform an updated prediction. Figure 1 (on page 56) shows a technical overview of the solution.

Now, at each of our hospitals, we are optimizing the solution in partnership with the care teams responsible for identifying and intervening on decompensating patients. We aim to reduce the rates of rapid response events, ICU transfers and ICU length of stay, and in-hospital mortality.

Here is a brief overview of the Adult Deterioration Solution's status at each hospital, including select news-worthy updates as well as how we are implementing them with our frontline care teams:

Duke University Hospital

- What is the status? The Adult Deterioration solution went live on May 28th, 2025 on all adult patient intermediate and stepdown units at Duke University Hospital. As of mid-April, the solution has generated 3,966 notifications on 1,437 unique patient hospitalizations. The solution sends 8.7 notifications per day to the Patient Response Team (PRT) Charge Nurse.
- How is the solution applied in the clinical workflow? Figure 2 shows the clinical workflow wherein the PRT nurse receives an automated notification. The notification is sent to the PRT charge nurse's mobile device as a page, using the Spok mobile application. The content of each notification reads as: "Adult Deterioration: ABC123 in 7110-01 is at HIGH risk of deterioration in the next 12 hours (calculated at 06:52AM 04/18)", where ABC123 is the patient's MRN, 7110-01 is bed, HIGH is the risk category, and 06:52AM is the time of prediction.
- Where are we headed? The PRT nursing team has helped us optimize the solution since its go live, including lengthening the snooze window from 4 to 6 hours, and excluding patients on comfort care. In partnership with this team and the Duke Heart Center, we are evaluating the added benefit of including clinical phenotype flags (Figure 3) for additional clinical context on why a patient may be deteriorating. We plan to include both the prediction model output and recent phenotype statuses in the content of each patient's page. We also plan to apply logic to risk thresholds and phenotype counts, optimizing for sensitivity and specificity of the notifications



Duke Raleigh Hospital

- What is the status? The solution went live at Duke Raleigh on January 6, 2026 on all adult patient intermediate and stepdown units at Duke Raleigh Hospital. By April 2026, the solution had generated 190 notifications on 73 unique patient hospitalizations. It sends 2.3 notifications per day to the Rounding Nurse Team.
- How is the solution applied in the clinical workflow? The Rounding Nurse Team follows a similar workflow to the PRT workflow at DUH, beginning with a notification formatted the same way as for DUH. Upon receiving a notification in Spok, the Rounding Nurse performs a chart review, then a brief bedside assessment. If the patient is actively deteriorating, the nurse calls an RRT event and huddles with the attending physician, looping in critical care as needed. Otherwise, if there is a deterioration concern, the nurse notifies the First Call physician and bedside nurse. In either case, the nurse adds the patient to their rounding list for the upcoming 24 hours and documents a note in Epic.
- Where are we headed? The Rounding Nurse Team Leads are collecting feedback from the nursing team and tracking paging volume versus note creation for the impacted patients. The goal is to understand workload and actionability of multiple pages on a given patient over time. The team is likewise interested in the inclusion of clinical phenotypes recently met for the page content, for context and improved performance related to identifying deterioration early.

Duke Regional Hospital

- What is the status? The Early Nurse Intervention Team (ENIT) leads are working with us to finalize the clinical workflow and assess impact on deteriorating patients in the intermediate/stepdown units at Duke Regional. From January through March 2026, the solution logged 119 notifications total across the five in-scope departments. We are comparing these notifications against ENIT consult orders placed, RRT events, Code Blues, ICU transfers, and in-hospital deaths. Based on this performance, and the workload capacity of the ENIT nurses, we will establish thresholds for CRITICAL, HIGH, and MEDIUM that are bespoke to Duke Regional.
- How is the solution applied in the clinical workflow? The ENIT Nurse workflow is similar to the workflows at DUH and DRAH: the ENIT nurse receives a page, performs a chart review and adds the patient to their prioritized rounding list, and performs a bedside assessment. If there is concern for deterioration, they engage the First Call Provider and Bedside
- Where are we headed? The Adult Deterioration Prediction solution is planned for go live at Duke Regional in Spring 2026. We will also plan to incorporate the clinical phenotypes in the notifications for additional clinical context and actionability.

We are collaborating with the Department of Population Health Sciences to study the impact of the solution on deterioration prevention, with the goal of publishing our results and further optimizing the predictive model performance. We are engaging with Duke Health Lake Norman now to understand their deterioration prevention workflows and investigate incorporation of the solution. We are also expanding into the Intensive Care Units across Duke Health to identify patients who are deteriorating within those care settings.

Notification Risk Category	PPV (12 hrs)	PPV (24 hrs)
Critical (n=333)	0.327	0.459
Critical or High (n=1,103)	0.260	0.369
Critical, High, or Medium (n=3,004)	0.170	0.266

Table 1: Adult Deterioration Positive Predictive Value Performance, prospective validation of calendar year 2024 on unanticipated ICU transfer (n = 1,370) or in-hospital death (n = 851)

Phenotype	Definition	Phenotype	Definition
Hypotension	Patient has either: 1) Two SBP < 90 mmHg within 6 hours 2) Two MAP < 65 mmHg within 6 hours	Vasoactive Medications	New administration or increased dose of at least one vasopressor or inotrope
End Organ Dysfunction	Patient has any of the following criteria: 1) Acute Renal Insufficiency (↑ Creatinine ≥ 0.3mg/dl or Creatinine ≥ 1.5x baseline) – exclusions: ESRD or active HD 2) Lactate > 2.0 mmol 3) AST or ALT > 5x ULN 4) Total Bilirubin > 2.0 mg/dL	Respiratory decline	Patient experiences any indicator of respiratory decline: 1) Increase in O2 (> 2L) within 6 hours 2) O2 saturation below 91% 3) PaO2 < 60mmHg 4) PaO2 decrease by 10mmHg
Hypoperfusion	Patient meets both: 1. Hypotension criteria AND 2. Develops End Organ Dysfunction criteria within 24 hours	Respiratory intervention	Patient has a new oxygen requirement (>6L via NC) or escalating level of respiratory support

Figure 1: Adult Deterioration Prediction solution technical overview: pulls data and calculates deterioration risk every 20 minutes

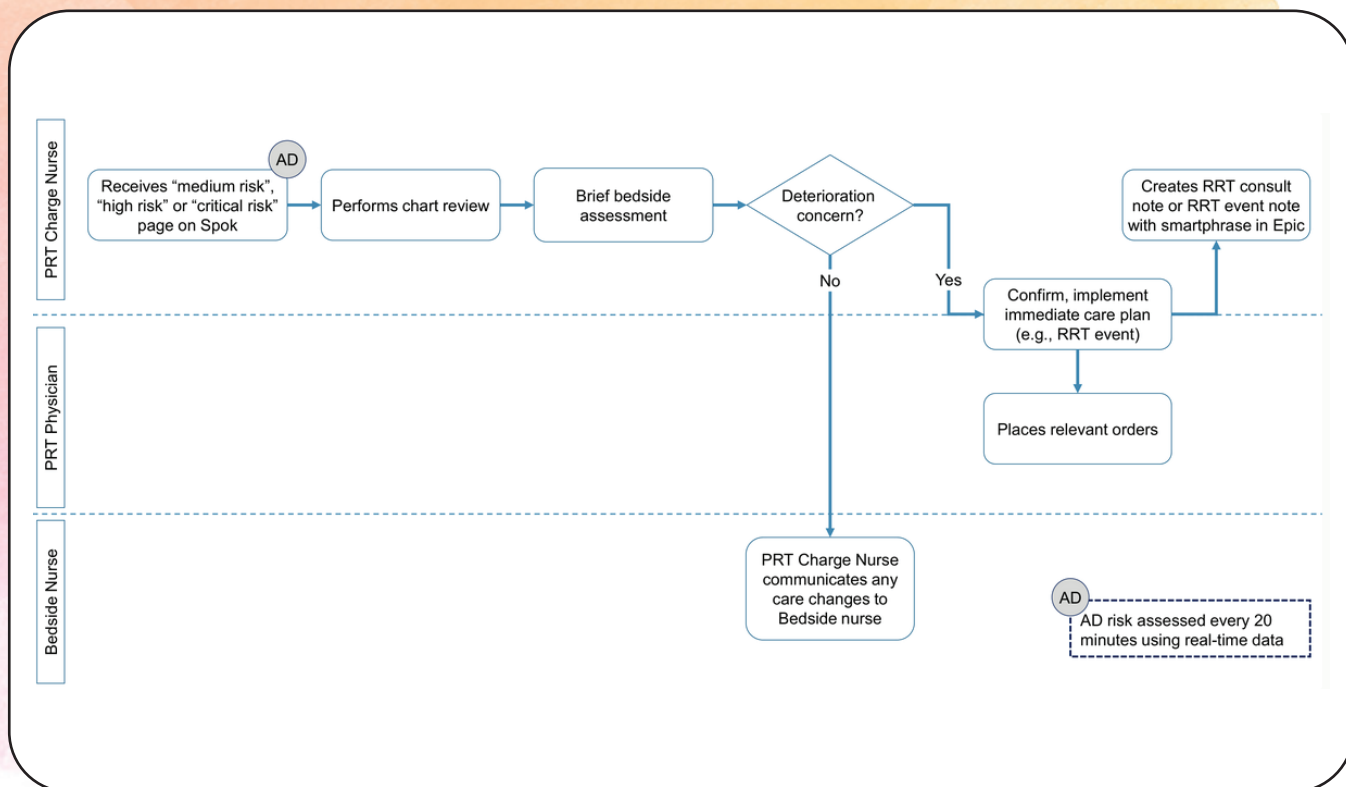


Figure 2: PRT workflow using the Adult Deterioration Prediction Tool at DUH

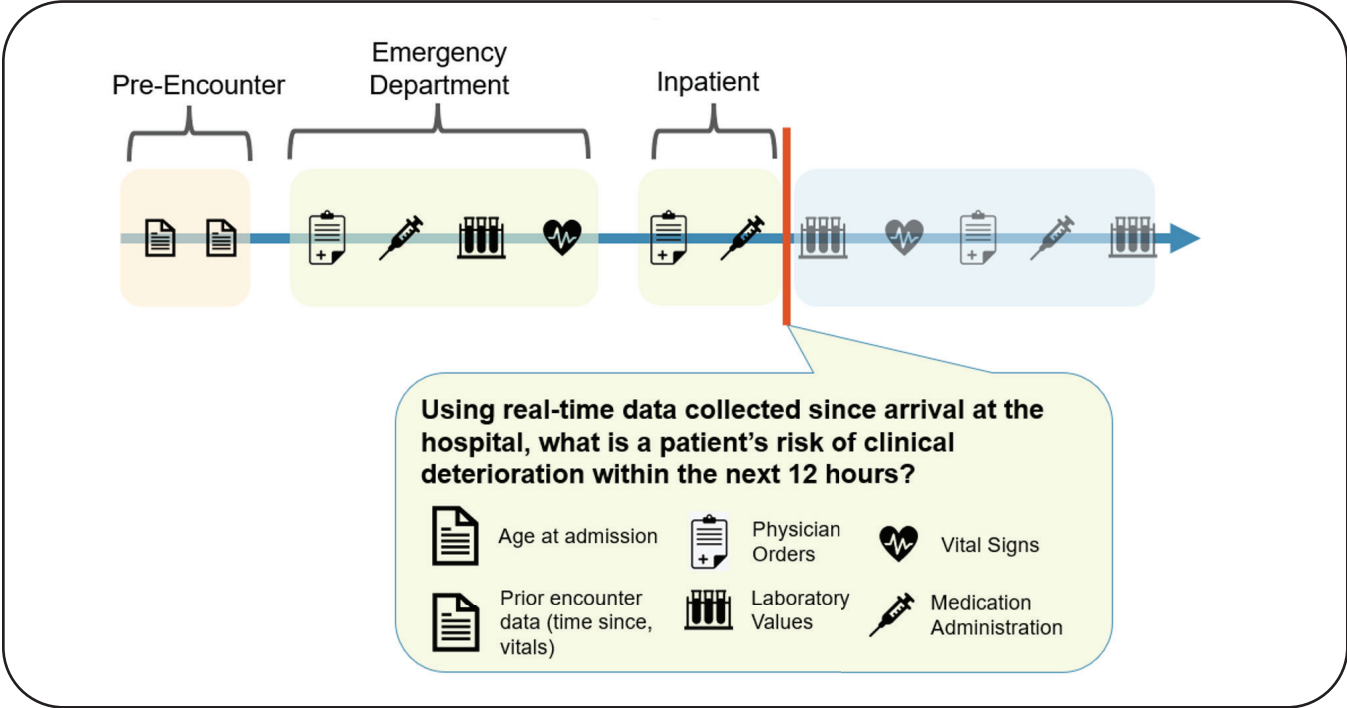


Figure 3: Clinical Phenotypes for Decompensation, based on the SCAI shock classification and iteration with Duke Heart Center

Department	Critical	Critical or High	Critical or High or Medium
DUH N8200 General Medicine	0.061	0.122	0.286
DUH N9200 Hematology Oncology	0.118	0.118	0.235
DUH N8300 General Surgery	0.043	0.100	0.229
DMP 8E General Medicine Stepdown	0.094	0.163	0.225
DCT 10A Adult BMT	0.021	0.062	0.188
DCT 10B Hematology Oncology	0.080	0.160	0.180
DUH N4300 General Medicine	0.043	0.103	0.154
DUH N7800 Renal/Pulmonary/MICU	0.056	0.098	0.154
DUH N9300 Oncology	0.024	0.065	0.121
DUH N2100 General Surgery	0.034	0.051	0.119

Table 2: Adult Deterioration Sensitivity Performance on top 10 units (50% of total outcomes), prospective validation of calendar year 2024 on unanticipated ICU transfer (n = 1,370) or in-hospital death (n = 851)



'CADEN' Real-Time AI Detection of Hospital-Acquired Conditions

James J. Jung, MD
Sai Samyukta Pale, MEng
Mike Revoir, BS
Marshall Nichols, MS
Will Knechtle, MBA, MPH
Matt Gardner, BS
Suresh Balu, MS, MBA

Project Brief

In high-acuity surgical environments, a persistent gap exists between what clinicians observe at the bedside and what gets formally captured in the medical record, with real consequences for patient care and hospital operations. The Clinical Condition Ascertainment & Detection Engine (CADEN) is a real-time AI platform built to close that gap, functioning as a clinical co-pilot embedded directly within existing clinician workflows.

Problem

The pace of care in surgical settings routinely outstrips the pace of formal documentation. Clinicians may recognize and act on a patient's condition in real time, but without structured documentation, that clinical reality remains invisible to the medical record and to every downstream system that depends on it.

This creates a compounding problem. An incomplete record undermines early deterioration detection, hinders care team coordination, and misrepresents the true complexity of the patient's condition. Under-documented patient acuity translates directly into a lower Case Mix Index (CMI), meaning hospital reimbursement falls short of the actual resources deployed.

Solution

When CADEN identifies a potential condition based on near real-time lab values and clinical data, it surfaces evidence-based documentation support within the clinician's existing workflow via dot phrases in native EHR note templates. No new interface. No additional administrative burden. Just timely, relevant prompts at the point of care.

CADEN was deployed on January 12, 2026, across six major surgical divisions at Duke: Urology, Transplant, Cardiothoracic, Vascular, Minimally Invasive, and General Surgery.

Impact

Diagnosis capture rates increased across all six targeted conditions compared to the equivalent pre-deployment period: Hyponatremia +126%, Hypokalemia +118%, Acute Blood Loss Anemia +41%, Hyperkalemia +27%, Hypernatremia +20%, Acute Kidney Injury +5%.

Next Steps

Near-term priorities include broadening CADEN's reach to additional hospital divisions and extending the target condition library. Ultimately, the vision is a medical record that is a true, high-fidelity reflection of every patient's clinical journey.

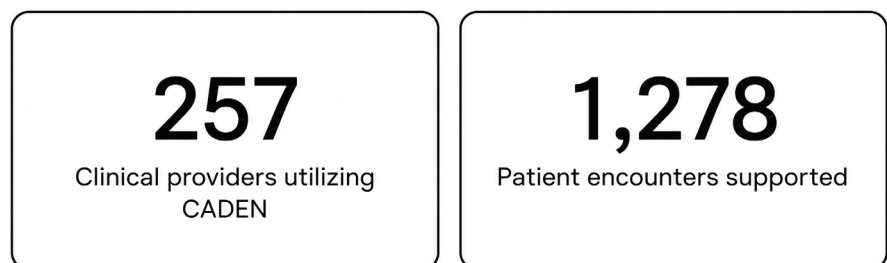


Figure 1: CADEN reach — 12-week deployment across 6 surgical divisions.



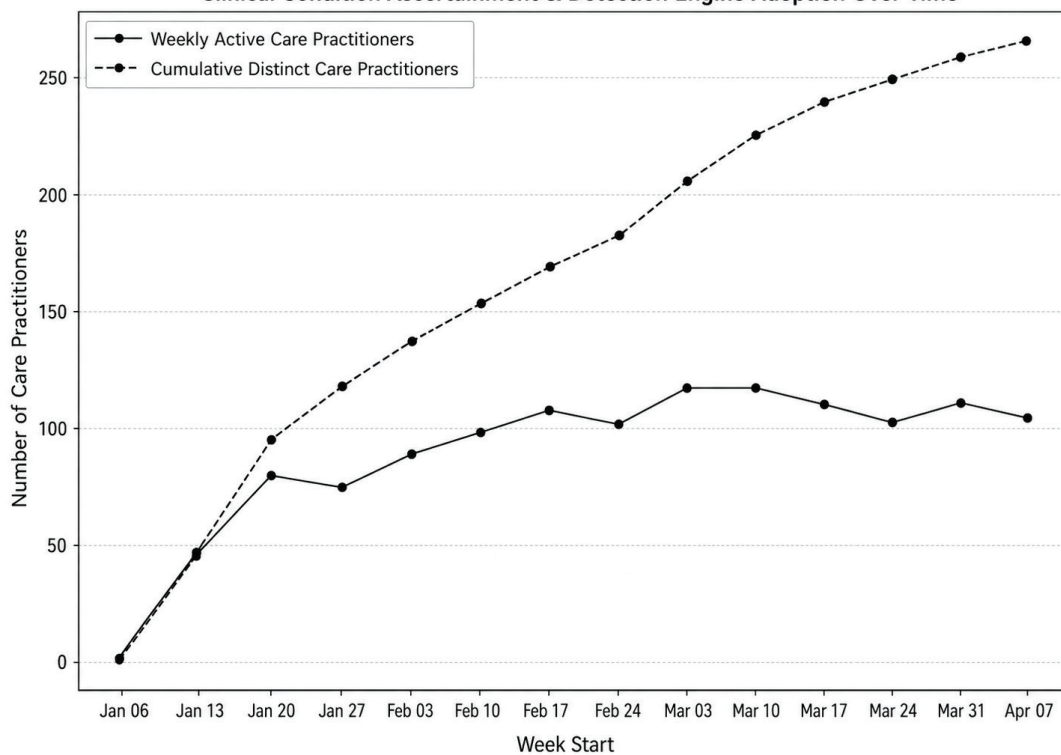


Figure 2: Increase in condition capture rate by condition type.

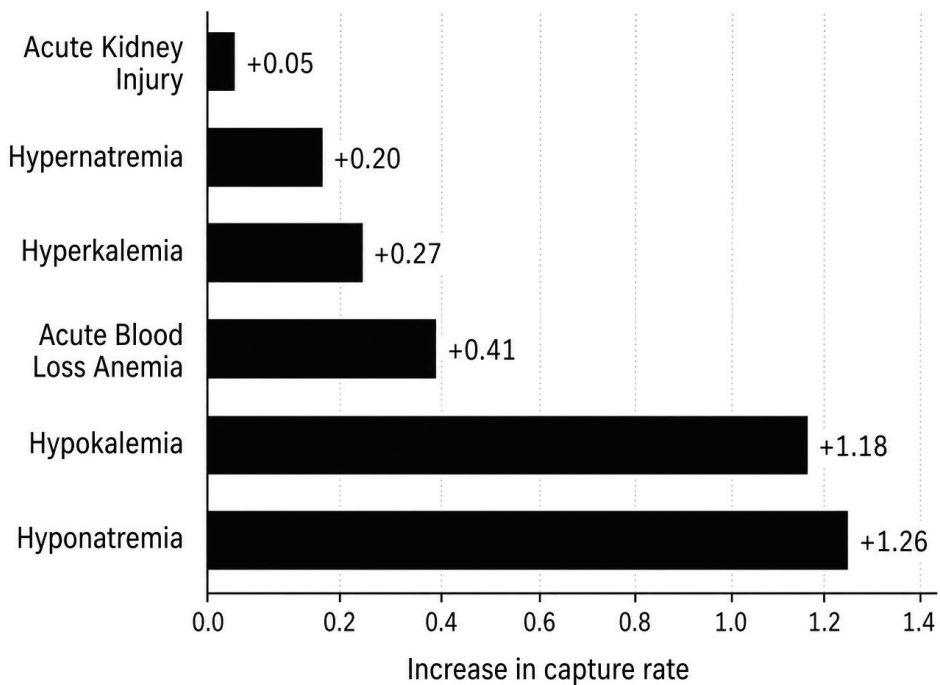


Figure 3: CADEN detection workflow overview — from real-time lab data to clinician notification and documentation support.

DIHI Analytics & AI for the DUHS 340B Drug Pricing Program

Marshall Nichols, MS
Will Knechtle, MBA, MPH
Suresh Balu, MS, MBA

The 340B Drug Pricing Program is a federal program allowing eligible hospitals and clinics (called Covered Entities) to purchase outpatient medications at significantly discounted prices. Duke University Health System (DUHS) participates through Duke University Hospital and Duke Regional Hospital. The financial benefit of the program is only realized when medication dispensing is properly linked to qualifying patient encounters. In practice, gaps exist: patients often receive prescriptions at ineligible DHIP locations after being seen at eligible DUH locations, and the third-party software DUHS relies on to match these encounters frequently misses valid claims.

DIHI was engaged to build tooling that could surface these missed linkages. Starting in early 2025, DIHI's data engineering team developed a 340B dashboard in both Tableau and Apache Superset. The dashboard queries Epic encounter and prescription data to identify cases where a patient had a qualifying visit at an eligible DUH location within 18 months of receiving a specialty prescription at an ineligible DHIP site.

Impact

The Duke Institute for Health Innovation 340B solution contribution has been separately and formally documented. DUHS Finance has formally reviewed the initiative, and it has been positioned as a recognized margin improvement contribution for the health system.





Publications & Collaborators



Publications

1. Fudim, Marat; Patel M; Balu S; et al. "CONFIDENT-HFpEF: A Machine Learning-Based Risk Stratification for Mortality and Hospitalization Using Multimodal Real-World Data." ESC Heart Failure (2026): xvag097.
2. Gao, Michael. From Bench to Bedside: Advancing the Evaluation and Alignment of Machine Learning and Large Language Models for Clinical Impact in Healthcare. Diss. Duke University, 2025.
3. Gao, Michael; Pejavara K; et al. "Development of a Flexible Chain of Thought Framework for Automated Routing of Patient Portal Messages." AMIA Annual Symposium Proceedings. Vol. 2024. 2025.
4. Gulamali, Freya; Kim JY; Pejavara K; Thomas C; et al. "Eliminating the AI digital divide by building local capacity." PLOS Digital Health 4.10 (2025): e0001026.
5. Hasan, Alifia; Prizant N; Kim JY; et al. "Aligning AI principles and healthcare delivery organization best practices to navigate the shifting regulatory landscape." NPJ Digital Medicine 8.1 (2025): 278.
6. Jackson, Brian R.; Sendak M; Balu S; et al. "Regulation of artificial intelligence in healthcare: Clinical Laboratory Improvement Amendments (CLIA) as a model." Journal of the American Medical Informatics Association 32.2 (2025): 404-407.
7. Kim, Jee Young, et al. "Invisible Infrastructure: Bridging the Digital Divide in AI Readiness for Safety-Net Healthcare Organizations." Available at SSRN 6183218 (2026).
8. Kim, Jee Young; Hasan A; Balu S; et al. "People process technology and operations framework for establishing AI governance in healthcare organizations." npj Digital Medicine (2026).
9. Kim, Jee Young; Hasan A; et al. "Establishing organizational AI governance in healthcare: a case study in Canada." NPJ Digital Medicine 8.1 (2025): 522.
10. Kpodzro, Sena, et al. "A Collaborative Best Practice Guide for Promoting AI Vendor Transparency in Health Care—The HAIP AI Vendor Disclosure Framework." NEJM AI 3.5 (2026): Alp2500985.



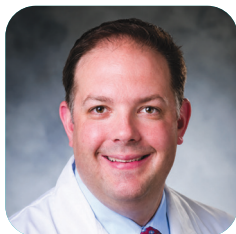
11. Lin, Anthony; Balu S.; et al. "Algorithm-Based Clinical Decision Support: Evolving Regulatory Landscape and Best Practices for Local Oversight." *Annual Review of Biomedical Data Science* 8 (2025).
12. Ma, Jessica E.; Gao, Michael; et al. "Impact of prognostic notifications on inpatient advance care planning: A cluster randomized trial." *Journal of Pain and Symptom Management* (2025).
13. Patel, Mihir N.; Gao, Michael; et al. "Physician Characteristics Associated With Advance Care Planning After a Machine Learning-Based Nudge." *American Journal of Hospice and Palliative Medicine®* (2025): 10499091251401462.
14. Prizant, Noah; Niederhofer K; Knechtle W; Hintze B; Gao M; Gardner M; et al. "Development, Validation, and Silent Trial of a Predictive Model for Acute Care Use in Cancer Patients on Immunotherapy." *AI in Precision Oncology* 2.5 (2025): 177-186.
15. Sendak, Mark; Kim JY; Hasan A; Balu S; et al. "Building models, building capacity: a review of participatory machine learning for HIV prevention." *PLOS Global Public Health* 5.6 (2025): e0003862.
16. Thomas, Ciera, et al. "AI Implementation in Safety Net Healthcare: Understanding Barriers and Strategies." *medRxiv* (2026): 2026-04.
17. Tyner-Monroe, Sam; Kim JY; Balu S; et al. "Understanding and Mitigating Unintended Bias in Medical AI Systems." *Harvard Data Science Review* (2026).
18. Valan, Bruno; Prakash A; Ratliff W; Gao, Michael; et al. "Evaluating Sepsis Watch generalizability through multisite external validation of a Sepsis Machine learning model." *npj Digital Medicine* 8.1 (2025): 350.
19. Young, Karen; Balu S; et al. "S2394 Metabolic Dysfunction-Associated Steatotic Liver Disease and Hormone Therapy for Gender-Affirming Care." *Official journal of the American College of Gastroenterology* | *ACG* 120.10S2 (2025): S515.



DIHI Innovation Core



Suresh Balu, MS, MBA
Executive Director for DIHI
Associate Dean,
Innovation and Partnership,
School of Medicine



Jeffrey Ferranti, MD, MS
Senior Vice President,
Chief Digital Officer
Duke University Health System



Jamie Daniel, BS
Solutions Architect



Michael Gao, PhD
Data Scientist



Matt Gardner, BS
Solutions Architect



Alifia Hasan, B.Pharm, MBA
Innovation Program Manager



Bradley Hintze, PhD
Data Engineer



Jee Young Kim, PhD
Qualitative Research Scientist



Will Knechtle, MBA, MPH
Innovation Program Manager



Sena Kpodzro, MPH
Research & Product Manager



Marshall Nichols, MS
Data Engineer



Kartik Pejavara, BS
Data Scientist



Samyukta Palle, MEng
Data Scientist



Anusha Prakash, MSc
Data Scientist



Will Ratliff, MBA
Innovation Program Manager



Mike Revoir, BS
Solutions Architect



Shems Saleh, MS
Data Scientist



Ciera Thomas, MPH
Research & Product Manager



Jiunn-Tyng Yeh, MD, PhD
Clinical Implementation Lead

Students & Trainees

Anika Mohan
Ethan Rodgers-Gates
Jay Swayambunathan, MD

Jonathan Revoir
Nicklaus Reed
Nikkie Dutta, MS

Srijan Bhasin, MD
Tareq Aljurf, MD
Victor Vladimirov

Collaborators

Suresh Agarwal, MD
 Ayman Ali, MD
 Peter Allen, MD
 Colette Allen, NP
 Megan Arguello, RN
 Randy Arvay, PhD
 Jon Bae, MD
 Andrew Barbas, MD
 Matthew Barber, MD, MHS
 Melissa Bauer, DO
 Armando Bedoya, MD, MCCi
 Andrew Berchuck, MD
 Joel Boggan, MD, MPH
 John Bonini, BS
 Brandi Bottiger, MD, MBI
 M. Alan Brookhart, PhD
 Tres Brown III, BS
 Dan Bruno, MS
 Brian Burrows, MD
 Blake Cameron, MD, MBI
 Erik Carvalho, PT, DPT, MHA
 David Casarett, MD, MA
 Anand Chowdhury, MD, MCCi
 Saumil Chudgar, MD, MSED
 Richard Chung, MD
 Katy Clark, RN
 David Claxton, MMCi
 Kathleen Cooney, MD
 Allison (Alice) Cooper, OGNP
 Daniel Costello, MPA
 Matt Cox, BS
 Crystal Crider, RN

Angela Cummings, MHA, MSN, RN
 Lesley Curtis, PhD
 Simon Curtis, MHA
 Dan Dardani, CLP
 Mike Datto, MD, PhD
 Marshall Davis-Henaine, MS, MPH
 Kristen Dicks, MD, MPH
 Dawn Dolby, RN
 Catie Dunn, BSN
 Matthew Ellis, MD
 Jill Engel, DNP
 Katia Ferguson, MSN
 Lisa Ferrari, MD
 Jonathan Fischer, MD
 Sam Fiske, BSN
 Katie Flanagan, MSW, LCSW, ACM-SW
 Emily Fuhr, BS
 William (Bill) Fulkerson, MD
 Charles Gerardo, MD
 Oren Gottfried, MD
 Bradi Granger, MSN, PhD
 Brian Griffith, MD, MMCi
 Barbara Griffith, MD
 Marcus Gumbrecht, PharmD
 Rajan Gupta, MD
 Mike Haglund, MD, PhD, MEd
 Bobette Haley, BS
 Erin Hamtil, RN
 Matthew Hartwig, MD
 John Haws, PhD
 Ricardo Henao, PhD
 Jacqueline Henson, MD, MMCi

Adrian Hernandez, MD
 James (Jim) Howard, MD, MPH
 Elizabeth Howe, MBOE
 Brenna Hughes, MD, MSc
 Tracy Hughes, MMCi
 Aatif Husain, MD
 Sara Johnson, MBA
 Aubrey Jolly Graham, MD
 W. Schuyler Jones, MD
 James Jung, MD, PhD
 Aparna Kamath, MD
 Neel Kapadia, MD
 Deborah Kaye, MD, MS
 Hadiza Kazaure, MD
 Carly Kelley, MD
 Michael Kent, MD
 Tara Kinard, RN, MSN, MBA
 Joanna Kipnes, MD, MS
 Allan Kirk, MD, PhD
 Stuart J Knechtle, MD
 Charley Kneifel, PhD
 Kasia Konopka, MFA
 Catherine Kuhn, MD
 Andrii Kuraksa, MBA
 Sandhya Lagoo-Deenadayalan, MD, PhD
 Kristen Lakis, Mdiv, LCSW
 David Leverenz, MD
 Jennifer Li, MD
 Cooper Linton, MHA, MBA
 Michael Lipkin, MD, MBA
 Kristen Lott, MSN, BSN
 Teri-Lynne Sennett



DUKE INSTITUTE FOR HEALTH INNOVATION

Kay Lytle, DNP
Mary Martin, MPH
Shelley McDonald, DO, PhD
Terry McDonnell, DNP
Carl McKinley, JD, MBA
Sachin Mehta, MD
Angelo Milazzo, MD, MBA
Cory Miller, RN
Susanna Naggie, MD, MPH
Aditee Narayan, MD
Emily Norboge
Kevin Oeffinger, MD
Kitty O'Hare, MD
Lance Okeke, MD
Paige Oldiges, MS
Karen Osborne, RN, BSN
Theodore Pappas, MD
Chet Patel, MD
Manesh Patel, MD
Milap Patel, MD
Juliessa Pavon, MD
Lisa Pickett, MD
Jessie Pinnix

Mihai Podgoreanu, MD
Amy Podraza, MBA
Eric Poon, MD, MPH
Judy Prewitt, RN
Robin Rasor, MS
Ann Reed, MD
Derek Robinson, MBA
Matt Roman, MHA, MMCI
Matt Rougeux, MHA
Moiria Rynn, MD
Mary Schilder, RN
Jessica Seidelman, MD
Mya Sendak, MD
Kevin Shah, MD, MBA
Colleen Shannon, JD
Rick Shannon, MD
Shilpa Shelton, MHA
Lauren Siewny, MD
Ryan Smith, MPP, MTS
S. Benjamin Smith, PhD
Sophia Smith, PhD
Becky Smith, MD
Susan Spratt, MD

Emily Sterrett, MD
Keith Stover, MAcc
Neel Subramanian, MD
Debra Sudan, MD
Aparna Swaminathan, MD
James Tcheng, MD
Jason Theiling, MD
Alison Toth, MD
Roya Tran, PharmD, MS
Chris Trumbo, RN
Sait Tunc, PhD
Hope Uronis, MD, MHS
Sreekanth Vemulapalli, MD
Lalit Verma, MD
Anthony Viera, MD, MPH
Ton Wang, MD, MS
Wendy Webster, MBA
Heidi White, MD, MHS, MEd, CMD
Sean Williams, MS
Michael Yarrington, MD, MCCi
Aimee K. Zaas, MD, MHS
Danyang Zhuo, PhD





Duke Institute for Health Innovation
www.dihi.org