Ensuring the Safe, Effective, and Equitable Translation of AI/ML Into Clinical Practice

Mark Sendak, MD, MPP
Population Health & Data Science Lead, Duke Institute for Health Innovation
Co-Lead, Health AI Partnership

Suresh Balu, MBA, MS
Director, Duke Institute for Health Innovation
Associate Dean for Innovation and Partnerships, Duke School of Medicine
Co-Lead, Health AI Partnership

January 2024
<table>
<thead>
<tr>
<th>Topic</th>
<th>Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duke Institute for Health Innovation</td>
<td>2 mins</td>
</tr>
<tr>
<td>Health AI Partnership</td>
<td>2 mins</td>
</tr>
<tr>
<td>Safe, Effective, and Equitable AI Translation</td>
<td>20 mins</td>
</tr>
<tr>
<td>Health Equity Across the AI Lifecycle (HEAAL)</td>
<td>8 mins</td>
</tr>
<tr>
<td>Duke Institute for Health Innovation</td>
<td>2 mins</td>
</tr>
<tr>
<td>---------------------------------------</td>
<td>--------</td>
</tr>
<tr>
<td>Health AI Partnership</td>
<td>2 mins</td>
</tr>
<tr>
<td>Safe, Effective, and Equitable AI Translation</td>
<td>20 mins</td>
</tr>
<tr>
<td>Health Equity Across the AI Lifecycle (HEAAL)</td>
<td>8 mins</td>
</tr>
</tbody>
</table>
Duke Institute for Health Innovation

Our Mission: **Catalyze innovations at Duke**
Catalyze transformative innovation in health and healthcare through high-impact research, leadership development and workforce training and the cultivation of a community of entrepreneurship

Our Approach: **Innovation by design**
Understand user workflow, desired outcomes and problems (needs) and then collaboratively develop concepts and prototypes, and iterate through to finalize solution
DIHI domains of innovation

Duke Institute for Health Innovation

Implementation and Health Delivery Science
- Catalyze multidisciplinary teamwork
- New care models
- Structured interface to Duke Health
- Living laboratory to incubate, refine, validate, and scale new ideas

Health Technology Innovation
- Leverage a growing health data infrastructure
- Create a connected digital health ecosystem
- Collaboration and co-development of technology
- Responsible development of datascience solutions

Leadership and Workforce Development
- Train current and future leaders across health care: Leadership Management Innovation Quantitative health sciences
- Contribute to developing the workforce of the future

Best Practices Development and Dissemination
- Disseminate best practices derived through internal R&D to elevate health innovation across ecosystem
- Convene stakeholders across settings to address common challenges in health innovation
DIHI Innovation Jam

A Health focused Shark Tank at Duke

- Solicits and identifies high-potential healthcare and health innovations ready for commercialization
- Duke Leadership as Sharks:
  - DUHS leaders, Department Chairs, Deans of School of Medicine, Nursing, Engineering, OLV, I&E, MedBlue, Center and Institute Directors
  - Innovation proposals from students, faculty, trainees and staff across campus
- Funding to support entrepreneurship / formation of company and also develop the product/service etc.
- Inventors offer portion of their share of Duke internal returns for investment from the sharks
- Internal syndicated investment agreements documented through MOUs.

DIHI RFA approach

“Top-down + Bottom-Up” approach to sourcing innovations

- Duke Health leadership develops mission-aligned strategic themes for innovation
- Front-line faculty and staff propose “problems” aligned with strategic themes and novel solutions
- Systematic review and due diligence: Assessments on team, feasibility, resource needs, impact and value to patients
- Operational Lead engaged right from the proposal stage
- 8-12 innovations funded each year; Duration: 12-15 months
- DIHI members embedded within project innovation teams to rapidly catalyze the innovations
- Pivots as needed to support rapid evolution to create value
- Metrics: clinical utility, economic utility, cultural impact, IP and academic outputs

<table>
<thead>
<tr>
<th>RFA</th>
<th>DIHI Innovation Jam</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;Top-down + Bottom-Up&quot; approach to sourcing innovations</td>
<td>A Health focused Shark Tank at Duke</td>
</tr>
<tr>
<td>Duke Health leadership develops mission-aligned strategic themes for innovation</td>
<td>Solicits and identifies high-potential healthcare and health innovations ready for commercialization</td>
</tr>
<tr>
<td>Front-line faculty and staff propose “problems” aligned with strategic themes and novel solutions</td>
<td>Duke Leadership as Sharks:</td>
</tr>
<tr>
<td>Systematic review and due diligence: Assessments on team, feasibility, resource needs, impact and value to patients</td>
<td>- DUHS leaders, Department Chairs, Deans of School of Medicine, Nursing, Engineering, OLV, I&amp;E, MedBlue, Center and Institute Directors</td>
</tr>
<tr>
<td>Operational Lead engaged right from the proposal stage</td>
<td>- Innovation proposals from students, faculty, trainees and staff across campus</td>
</tr>
<tr>
<td>8-12 innovations funded each year; Duration: 12-15 months</td>
<td>- Funding to support entrepreneurship / formation of company and also develop the product/service etc.</td>
</tr>
<tr>
<td>DIHI members embedded within project innovation teams to rapidly catalyze the innovations</td>
<td>- Inventors offer portion of their share of Duke internal returns for investment from the sharks</td>
</tr>
<tr>
<td>Pivots as needed to support rapid evolution to create value</td>
<td>- Internal syndicated investment agreements documented through MOUs.</td>
</tr>
<tr>
<td>Metrics: clinical utility, economic utility, cultural impact, IP and academic outputs</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Unstructured</th>
<th>Structured</th>
</tr>
</thead>
<tbody>
<tr>
<td>DIHI RFA approach</td>
<td>DIHI Innovation Jam</td>
</tr>
<tr>
<td>&quot;Top-down + Bottom-Up&quot; approach to sourcing innovations</td>
<td>A Health focused Shark Tank at Duke</td>
</tr>
<tr>
<td>Duke Health leadership develops mission-aligned strategic themes for innovation</td>
<td>Solicits and identifies high-potential healthcare and health innovations ready for commercialization</td>
</tr>
<tr>
<td>Front-line faculty and staff propose “problems” aligned with strategic themes and novel solutions</td>
<td>Duke Leadership as Sharks:</td>
</tr>
<tr>
<td>Systematic review and due diligence: Assessments on team, feasibility, resource needs, impact and value to patients</td>
<td>- DUHS leaders, Department Chairs, Deans of School of Medicine, Nursing, Engineering, OLV, I&amp;E, MedBlue, Center and Institute Directors</td>
</tr>
<tr>
<td>Operational Lead engaged right from the proposal stage</td>
<td>- Innovation proposals from students, faculty, trainees and staff across campus</td>
</tr>
<tr>
<td>8-12 innovations funded each year; Duration: 12-15 months</td>
<td>- Funding to support entrepreneurship / formation of company and also develop the product/service etc.</td>
</tr>
<tr>
<td>DIHI members embedded within project innovation teams to rapidly catalyze the innovations</td>
<td>- Inventors offer portion of their share of Duke internal returns for investment from the sharks</td>
</tr>
<tr>
<td>Pivots as needed to support rapid evolution to create value</td>
<td>- Internal syndicated investment agreements documented through MOUs.</td>
</tr>
<tr>
<td>Metrics: clinical utility, economic utility, cultural impact, IP and academic outputs</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>11 Years</th>
<th>90+ Innovation Projects</th>
<th>740+ Proposals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Catalyzing Innovations</td>
<td>Innovation Projects</td>
<td>Proposals</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>6 Years</th>
<th>30+ Pitches</th>
<th>12 Companies</th>
</tr>
</thead>
<tbody>
<tr>
<td>of Jamming</td>
<td>Incubated</td>
<td></td>
</tr>
</tbody>
</table>

Industry best-practice approach in catalyzing innovation
We invite you to submit your novel ideas supporting Generative AI & Large Language Models: AI solutions to improve staff and clinician efficiency, patient journey and outcomes.

Visit: dihi.org/events/rfa
email: dihi-rfa@duke.edu

Proposals due: NOVEMBER 3, 2023
DIHI Spectrum of Value Creation

Inpatient Innovations
- HIV Pre-Exposure Prophylaxis Identification
- CKD Patient Education Dissemination
- Community-Based Palliative Care
- High Value Analyte Ordering
- NAFLD population health rounding
- CKD population health rounding

Transition Setting
- High-utilizer dashboard
- Complex Care Plans
- Index Admissions with MSSP
- Readmissions (Social Drivers for HF)
- SNF transition
- SNF transition
- Index Admissions with MSSP
- Readmissions (Social Drivers for HF)
- SNF transition
- High-utilizer dashboard
- Complex Care Plans
- Index Admissions with MSSP
- Readmissions (Social Drivers for HF)
- SNF transition

Outpatient/Gaps in Care
- Community COVID-19 Support
- Patient Reported Outcomes for Cancer Patients
- Outpatient Procedure Concierge Program
- Cancer Distress Coach
- Autism and Beyond
- Voices of Duke

Patient & Community
- Immersion in innovation and data science
- Medical Students Scholarship
- Data Science in Health masters course in BME
- Summer Fellowship in Data Science
- Case Studies and Data Camp
- Journal Club

Technology Infrastructure
Research and Dissemination
Education and Training

Duke Institute for Health Innovation [DIHI] – Spectrum of value creation across the ecosystem
<table>
<thead>
<tr>
<th>Topic</th>
<th>Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duke Institute for Health Innovation</td>
<td>2 mins</td>
</tr>
<tr>
<td>Health AI Partnership</td>
<td>2 mins</td>
</tr>
<tr>
<td>Safe, Effective, and Equitable AI Translation</td>
<td>20 mins</td>
</tr>
<tr>
<td>Health Equity Across the AI Lifecycle (HEAAL)</td>
<td>8 mins</td>
</tr>
</tbody>
</table>
Corps Sites

Website healthaipartnership.org
Our Mission: Empowering healthcare professionals to use AI effectively, safely, and equitably through community-informed up-to-date standards

Our Values

- **advance health equity**
  - prioritize solutions that advance health equity and eliminate the AI digital divide

- **improve patient care**
  - ensure that AI adoption is driven by patient care needs, not technical novelty

- **improve the workplace**
  - surface socio-technical challenges in AI use and foster a positive work environment

- **build community**
  - create safe spaces to share learnings and consult peers
Phase One (Apr 22 – Aug 23) Milestones

**Standard AI Solution Procurement Milestones**

- Community-informed best practices sourced from across the network of organizations
- Multiple co-design workshops with IDEO.org
- Focused on AI solutions used for:
  - Diagnosis or treatment decisions for individual patients
  - Prioritization of patients for healthcare services (e.g., surgery scheduling, care management prioritization, ED triaging)

**Health Equity Across the AI Lifecycle (HEAAL) Framework**

- Developed to answer the question: “our health system is considering adopting a new solution that uses AI; how do we assess the potential future impact on health inequities?”
- Convened multi-stakeholder workshop featuring case studies, expert discussants, and framework developers
- Developed detailed procedures for healthcare organizations to follow for AI procurement

**Key Decision Points**
- 8

**Interviews**
- 85+

**Topic Guides**
- 31

**Case Studies**
- 3

**Participants**
- 75+

**Procedures**
- 37
<table>
<thead>
<tr>
<th>Topic</th>
<th>Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duke Institute for Health Innovation</td>
<td>2 mins</td>
</tr>
<tr>
<td>Health AI Partnership</td>
<td>2 mins</td>
</tr>
<tr>
<td>Safe, Effective, and Equitable AI Translation</td>
<td>20 mins</td>
</tr>
<tr>
<td>Health Equity Across the AI Lifecycle (HEAAL)</td>
<td>8 mins</td>
</tr>
</tbody>
</table>
8 Key Decision Points in AI Adoption Process

1. Identify and prioritize a problem
2. Evaluate AI as a viable component of the solution
3. Develop measures of outcomes and success of the AI product
4. Design a new optimal workflow to facilitate integration
5. Evaluate pre-integration safety and effectiveness of the AI product
6. Execute change management, workflow integration, and scaling strategy
7. Monitor and maintain the AI product
8. Update or decommission the AI product
8 Key Decision Points in AI Adoption Process

1. Identify and prioritize a problem
2. Evaluate AI as a viable component of the solution
3. Develop measures of outcomes and success of the AI product
4. Design a new optimal workflow to facilitate integration
5. Evaluate pre-integration safety and effectiveness of the AI product
6. Execute change management, workflow integration, and scaling strategy
7. Monitor and maintain the AI product
8. Update or decommission the AI product
Align Front-Line Staff and Organizational Leaders

Create Alignment Throughout Project Selection

<table>
<thead>
<tr>
<th>Strategic Priorities</th>
<th>Proposal Solicitation</th>
<th>Project Selection</th>
<th>Scoping</th>
<th>Solution Development</th>
<th>Implementation &amp; Evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specify 4-5 organizational innovation priorities</td>
<td>Publicize RFA</td>
<td>Rate oral pitches to select ~10 projects</td>
<td>Define problem, solution, stakeholders, metrics, and measures</td>
<td>Design and develop solution with DIHI</td>
<td>Evaluate and determine impact and sustainability plans</td>
</tr>
</tbody>
</table>

C-Suite
- Specify 4-5 organizational innovation priorities
- Publicize RFA

Clinical and Operational Leaders (control resources)
- Publicize RFA and work with front-line staff on proposal
- Provide written reviews to select ~20 projects for oral pitches

Front-Line Staff
- Iterate on proposal with buy-in from clinical and operational leaders
- Define operationalization and dissemination, including communication and training
Align Front-Line Staff and Organizational Leaders

Create Alignment Throughout Project Selection

<table>
<thead>
<tr>
<th>Strategic Priorities</th>
<th>Proposal Solicitation</th>
<th>Project Selection</th>
<th>Scoping</th>
<th>Solution Development</th>
<th>Implementation &amp; Evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specify 4-5</td>
<td>Publicize RFA</td>
<td>Rate and</td>
<td>Evaluate and determine impact and sustainability plans</td>
<td></td>
<td></td>
</tr>
<tr>
<td>organizational</td>
<td></td>
<td>Communication</td>
<td></td>
<td>plans</td>
<td></td>
</tr>
<tr>
<td>innovation priorities</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

C-Suite

Specify 4-5 organizational innovation priorities

Publicize RFA

Publicize RFA and work with front-line staff on proposal

Iterate on proposal with buy-in from clinical and operational leaders

Rate and communicate

Communication

Iteration

Clinical and Operational Leaders (control resources)

Define problem, solution, stakeholders, metrics, and measures

Define operationalization and dissemination, including communication and training

Front-Line Staff

Evaluate and determine impact and sustainability plans

Design and develop solution with DIHI

Proprietary and Confidential
8 Key Decision Points in AI Adoption Process

1. Identify and prioritize a problem
2. Evaluate AI as a viable component of the solution
3. Develop measures of outcomes and success of the AI product
4. Design a new optimal workflow to facilitate integration
5. Evaluate pre-integration safety and effectiveness of the AI product
6. Execute change management, workflow integration, and scaling strategy
7. Monitor and maintain the AI product
8. Update or decommission the AI product
ML Data Quality Assurance for Healthcare

https://proceedings.mlr.press/v182/sendak22a.html
## Development and Validation of ML-DQA

<table>
<thead>
<tr>
<th>Phase I: Data Element Pre-Processing</th>
<th>Pediatric Sepsis Prediction</th>
<th>Lung Transplant Complication Prediction</th>
<th>Sepsis Prediction at Jefferson Health</th>
<th>Immune-Related Adverse Event Prediction</th>
<th>Maternal Morbidity and Mortality Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-existing groupers</td>
<td>108</td>
<td>109</td>
<td>30</td>
<td>39</td>
<td>310</td>
</tr>
<tr>
<td>Project-specific groupers</td>
<td>73</td>
<td>35</td>
<td>59</td>
<td>41</td>
<td>12</td>
</tr>
</tbody>
</table>

### Phase II: ML-DQA Checks

<table>
<thead>
<tr>
<th></th>
<th>Pediatric Sepsis Prediction</th>
<th>Lung Transplant Complication Prediction</th>
<th>Sepsis Prediction at Jefferson Health</th>
<th>Immune-Related Adverse Event Prediction</th>
<th>Maternal Morbidity and Mortality Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Completeness checks</td>
<td>144</td>
<td>144</td>
<td>70</td>
<td>508</td>
<td>404</td>
</tr>
<tr>
<td>Conformance checks</td>
<td>122</td>
<td>144</td>
<td>132</td>
<td>225</td>
<td>69</td>
</tr>
<tr>
<td>Plausibility checks</td>
<td>123</td>
<td>144</td>
<td>61</td>
<td>301</td>
<td>404</td>
</tr>
<tr>
<td>Total quality checks</td>
<td>389</td>
<td>432</td>
<td>267</td>
<td>1,034</td>
<td>877</td>
</tr>
</tbody>
</table>

Grouper Maintenance to Address Meta Data Instability

Which Creatinine?

Which Glucose?
8 Key Decision Points in AI Adoption Process

1. Identify and prioritize a problem
2. Evaluate AI as a viable component of the solution
3. Develop measures of outcomes and success of the AI product
4. Design a new optimal workflow to facilitate integration
5. Evaluate pre-integration safety and effectiveness of the AI product
6. Execute change management, workflow integration, and scaling strategy
7. Monitor and maintain the AI product
8. Update or decommission the AI product
## Example Categories of Measures

<table>
<thead>
<tr>
<th>Category</th>
<th>Definition</th>
<th>Example Metrics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model performance</td>
<td>Effectiveness, accuracy, and reliability of the AI model or algorithm in fulfilling its intended tasks within the clinical or healthcare context.</td>
<td>Sensitivity (recall, true positive rate), Specificity (true negative rate), Area Under the ROC Curve (AUC-ROC), F1 Score, Precision (positive predictive value).</td>
</tr>
<tr>
<td>Software performance</td>
<td>Efficiency and responsiveness of processing tasks, delivering results, and overall performance of the software components and its interactions.</td>
<td>Inference time, throughput, model latency, response time, resource utilization, scalability.</td>
</tr>
<tr>
<td>Clinical effectiveness</td>
<td>Assessment of impact of product use on healthcare outcomes.</td>
<td>Mortality rate, intensive care unit requirement, complication rate</td>
</tr>
<tr>
<td>Usability</td>
<td>Quality of users’ interactions with the AI-based medical software.</td>
<td>Clinician satisfaction, user error rates, ease of use.</td>
</tr>
<tr>
<td>Safety and security</td>
<td>Safe and secure operating software, evaluating harm to patients and protection against unauthorized access, data breaches, and cyber threats.</td>
<td>Number of identified safety risks and mitigations, adherence to cybersecurity standards, detection of adversarial attacks, incident response time.</td>
</tr>
<tr>
<td>Business</td>
<td>Business objectives and outcomes</td>
<td>Reduction in diagnostic time, cost savings.</td>
</tr>
</tbody>
</table>
Mortality Model Performance Measures

Table 2. Prediction Accuracy by Evaluation Method, Location, and Time

<table>
<thead>
<tr>
<th>Evaluation Method</th>
<th>Location</th>
<th>Time</th>
<th>AUROC (95% CI)</th>
<th>AUPRC (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Retrospective</td>
<td>Hospital A</td>
<td>2014-2015</td>
<td>0.87 (0.83-0.89)</td>
<td>0.29 (0.25-0.37)</td>
</tr>
<tr>
<td>Retrospective</td>
<td>Hospital A</td>
<td>2018</td>
<td>0.85 (0.83-0.87)</td>
<td>0.17 (0.13-0.22)</td>
</tr>
<tr>
<td>Retrospective</td>
<td>Hospital B</td>
<td>2018</td>
<td>0.89 (0.86-0.92)</td>
<td>0.22 (0.14-0.31)</td>
</tr>
<tr>
<td>Retrospective</td>
<td>Hospital C</td>
<td>2018</td>
<td>0.84 (0.80-0.89)</td>
<td>0.13 (0.08-0.21)</td>
</tr>
<tr>
<td>Prospective</td>
<td>Hospital A</td>
<td>2019</td>
<td>0.86 (0.83-0.90)</td>
<td>0.14 (0.09-0.21)</td>
</tr>
</tbody>
</table>

A. ROC curve

B. PR curve
# Mortality Model Performance Measures

<table>
<thead>
<tr>
<th>Threshold</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>PPV</th>
<th>Alerts, No./d</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>0.88</td>
<td>0.66</td>
<td>0.05</td>
<td>39.9</td>
<td>37.8</td>
<td>2.1</td>
<td></td>
</tr>
<tr>
<td>0.02</td>
<td>0.76</td>
<td>0.81</td>
<td>0.08</td>
<td>23.3</td>
<td>21.5</td>
<td>1.8</td>
<td></td>
</tr>
<tr>
<td>0.03</td>
<td>0.68</td>
<td>0.88</td>
<td>0.11</td>
<td>15.3</td>
<td>13.6</td>
<td>1.7</td>
<td></td>
</tr>
<tr>
<td>0.04</td>
<td>0.61</td>
<td>0.91</td>
<td>0.12</td>
<td>11.9</td>
<td>10.4</td>
<td>1.5</td>
<td></td>
</tr>
<tr>
<td>0.05</td>
<td>0.57</td>
<td>0.93</td>
<td>0.15</td>
<td>9.1</td>
<td>7.7</td>
<td>1.4</td>
<td></td>
</tr>
<tr>
<td>0.06</td>
<td>0.54</td>
<td>0.95</td>
<td>0.18</td>
<td>7.4</td>
<td>6.1</td>
<td>1.3</td>
<td></td>
</tr>
<tr>
<td>0.07</td>
<td>0.52</td>
<td>0.95</td>
<td>0.19</td>
<td>6.5</td>
<td>5.3</td>
<td>1.3</td>
<td></td>
</tr>
<tr>
<td>0.08</td>
<td>0.50</td>
<td>0.96</td>
<td>0.21</td>
<td>5.8</td>
<td>4.5</td>
<td>1.2</td>
<td></td>
</tr>
<tr>
<td>0.09</td>
<td>0.48</td>
<td>0.96</td>
<td>0.22</td>
<td>5.2</td>
<td>4.1</td>
<td>1.2</td>
<td></td>
</tr>
<tr>
<td>0.10</td>
<td>0.44</td>
<td>0.97</td>
<td>0.22</td>
<td>4.8</td>
<td>3.7</td>
<td>1.1</td>
<td></td>
</tr>
<tr>
<td>0.11</td>
<td>0.43</td>
<td>0.97</td>
<td>0.24</td>
<td>4.4</td>
<td>3.4</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>0.12</td>
<td>0.41</td>
<td>0.97</td>
<td>0.24</td>
<td>4.1</td>
<td>3.1</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>0.13</td>
<td>0.39</td>
<td>0.98</td>
<td>0.26</td>
<td>3.7</td>
<td>2.7</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>0.14</td>
<td>0.39</td>
<td>0.98</td>
<td>0.27</td>
<td>3.5</td>
<td>2.6</td>
<td>0.9</td>
<td></td>
</tr>
<tr>
<td>0.15</td>
<td>0.36</td>
<td>0.98</td>
<td>0.27</td>
<td>3.2</td>
<td>2.3</td>
<td>0.9</td>
<td></td>
</tr>
<tr>
<td>0.16</td>
<td>0.35</td>
<td>0.98</td>
<td>0.28</td>
<td>3.1</td>
<td>2.2</td>
<td>0.9</td>
<td></td>
</tr>
<tr>
<td>0.17</td>
<td>0.34</td>
<td>0.98</td>
<td>0.30</td>
<td>2.8</td>
<td>2.0</td>
<td>0.8</td>
<td></td>
</tr>
<tr>
<td>0.18</td>
<td>0.33</td>
<td>0.98</td>
<td>0.32</td>
<td>2.6</td>
<td>1.7</td>
<td>0.8</td>
<td></td>
</tr>
<tr>
<td>0.19</td>
<td>0.31</td>
<td>0.99</td>
<td>0.32</td>
<td>2.4</td>
<td>1.6</td>
<td>0.8</td>
<td></td>
</tr>
<tr>
<td>0.20</td>
<td>0.29</td>
<td>0.99</td>
<td>0.33</td>
<td>2.2</td>
<td>1.5</td>
<td>0.7</td>
<td></td>
</tr>
<tr>
<td>0.21</td>
<td>0.28</td>
<td>0.99</td>
<td>0.33</td>
<td>2.0</td>
<td>1.4</td>
<td>0.7</td>
<td></td>
</tr>
<tr>
<td>0.22</td>
<td>0.28</td>
<td>0.99</td>
<td>0.35</td>
<td>1.9</td>
<td>1.3</td>
<td>0.7</td>
<td></td>
</tr>
<tr>
<td>0.23</td>
<td>0.27</td>
<td>0.99</td>
<td>0.36</td>
<td>1.8</td>
<td>1.2</td>
<td>0.7</td>
<td></td>
</tr>
<tr>
<td>0.24</td>
<td>0.26</td>
<td>0.99</td>
<td>0.38</td>
<td>1.7</td>
<td>1.1</td>
<td>0.6</td>
<td></td>
</tr>
<tr>
<td>0.25</td>
<td>0.26</td>
<td>0.99</td>
<td>0.41</td>
<td>1.5</td>
<td>0.9</td>
<td>0.6</td>
<td></td>
</tr>
</tbody>
</table>

Abbreviation: PPV, positive predictive value.
8 Key Decision Points in AI Adoption Process

1. Identify and prioritize a problem
2. Evaluate AI as a viable component of the solution
3. Develop measures of outcomes and success of the AI product
4. Design a new optimal workflow to facilitate integration
5. Evaluate pre-integration safety and effectiveness of the AI product
6. Execute change management, workflow integration, and scaling strategy
7. Monitor and maintain the AI product
8. Update or decommission the AI product
Nephrology consult recommendation
Nephrology consult recommendation
Nephrology consult recommendation
Nephrology consult recommendation
Nephrology consult recommendation
Nephrology consult recommendation
Nephrology consult recommendation
Nephrology consult recommendation
Nephrology consult recommendation

Credit: RJ Andrews
Nephrology consult recommendation
Nephrology consult recommendation
“Doc, why didn’t anyone tell me sooner?”
Validated Measures

5 Year Risk of ESRD Progression - JAMA, 2011

2 Year eGFR Change - JAMA, 2014
Adapt Workflows, Roles, and Organization

Don’t Rely on Existing Workflows to Solve Problems

<table>
<thead>
<tr>
<th>Already seen Nephrologist</th>
<th>1068</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deceased</td>
<td>246</td>
</tr>
<tr>
<td>Don’t have CKD</td>
<td>83</td>
</tr>
<tr>
<td>ESRD</td>
<td>65</td>
</tr>
<tr>
<td>Eligible for Rounding</td>
<td>413</td>
</tr>
</tbody>
</table>

n=1875

n=46143
Adapt Workflows, Roles, and Organization

Don’t Rely on Existing Workflows to Solve Problems

1. Patient arrives with history of treatment from a variety of settings (at and outside of Duke).

2. All relevant patient data is aggregated and analyzed using algorithms and models that incorporate the best statistics theory and medical expertise.

3. An interdisciplinary team discusses the best plan. Team typically includes a specialists, PCP, data analyst, pharmacist, social worker, and care manager.

4. Next, an action plan is implemented. As new data becomes available, the evaluation and action plan process restarts.

Result: a better-coordinated, data-supported patient care
Adapt Workflows, Roles, and Organization

Don’t Rely on Existing Workflows to Solve Problems

<table>
<thead>
<tr>
<th>Intervention</th>
<th>N</th>
<th>% of Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nephrology Appointment</td>
<td>84</td>
<td>0.72</td>
</tr>
<tr>
<td>PCP Appointment</td>
<td>21</td>
<td>0.18</td>
</tr>
<tr>
<td>Lab Order</td>
<td>15</td>
<td>0.13</td>
</tr>
<tr>
<td>Medication Change</td>
<td>10</td>
<td>0.09</td>
</tr>
<tr>
<td>Care Management</td>
<td>7</td>
<td>0.05</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>137</td>
<td></td>
</tr>
</tbody>
</table>
Adapt Workflows, Roles, and Organization

Don’t Rely on Existing Workflows to Solve Problems

Now extended and applied to:
- Non-alcoholic fatty liver disease (NAFLD)
- Peripheral artery disease
- Community-based palliative care
Adapt Workflows, Roles, and Organization

Don’t Rely on Existing Workflows to Solve Problems

Now extended and applied to:
- Non-alcoholic fatty liver disease (NAFLD)
- Peripheral artery disease
- Community-based palliative care

“The difference in [algorithm] performance is negligible compared to the difference that a good physician champion makes, or a good intervention plan makes. Those are by far and away the most important things to the success of a project. The actual model itself is, as much as I might delude myself or whatever, it’s actually not that important.”

- Technical Stakeholder
Adapt Workflows, Roles, and Organization

Restructure Organization to Create Alignment

• Duke - Moved Rapid Response Team out of Cardiac ICU to create Patient Response Program with new reporting structure
• Duke - Moved care management function and ACO under newly created Population Health Management Office
• NYC – Moved Test + Trace out of Public Health Department directly into City Hall
8 Key Decision Points in AI Adoption Process

1. Identify and prioritize a problem
2. Evaluate AI as a viable component of the solution
3. Develop measures of outcomes and success of the AI product
4. Design a new optimal workflow to facilitate integration
5. Evaluate pre-integration safety and effectiveness of the AI product
6. Execute change management, workflow integration, and scaling strategy
7. Monitor and maintain the AI product
8. Update or decommission the AI product
Identifying Label Leakage During a Silent Trial

• Pediatric sepsis prediction
  – Outcome definition: Blood Culture ∩ Antibiotics for 4 days ∩ Acute organ dysfunction
  – LSTM with 6-hour prediction window and 3-hour snooze
  – Retrospective training set: 17,491 unique encounters for children between 30 days old and 18 years old between November 1, 2016 – December 31, 2020
  – Temporal validation set: 6,545 unique encounters for children between 30 days old and 18 years old between January 1, 2021 – June 30, 2022
Identifying Label Leakage During a Silent Trial

<table>
<thead>
<tr>
<th></th>
<th>AUROC</th>
<th>AUPRC</th>
<th>PPV at 20% sensitivity (with 3hr snooze)</th>
<th>PPV at 50% sensitivity (with 3hr snooze)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Retrospective test set</td>
<td>0.816</td>
<td>0.483</td>
<td>0.769</td>
<td>0.612</td>
</tr>
<tr>
<td>Temporal validation</td>
<td>0.862</td>
<td>0.386</td>
<td>0.851</td>
<td>0.611</td>
</tr>
</tbody>
</table>

**Silent Trial Design**

- A custom-built database extracted real-time patient data from EPIC every 15 minutes.
- The model calculated risk scores for all current encounters in the hospital.
- High risk notifications were sent to an internal HiPAA-compliant message channel.
- Alarm volumes were tracked and technical issues were resolved.
Identifying Label Leakage During a Silent Trial

• Silent trial results
  – Model ran on 1,475 unique encounters over 2 months
  – Model generated 30 alarms per day >> 2 alarms per day expected
  – Model fired alarm on almost all patients in ED within first hour of arrival
Identifying Label Leakage During a Silent Trial

• Label leakage due to layer normalization in LSTM
  – In retrospective training data:
    • set maximum encounter length to 168 hours
    • truncated sepsis encounters at time of sepsis
  – Shorter encounter → more padding of encounter hours with 0s
    → smaller mean after layer normalization
  – Longer encounter → less padding of encounter hours with 0s
    → larger mean after layer normalization
  – In retrospective data, model learned to associate early hours of encounter with sepsis
Identifying Label Leakage During a Silent Trial

- Retrained LSTM without layer normalization using the same hyperparameters

<table>
<thead>
<tr>
<th>Test Set and Normalization</th>
<th>AUROC</th>
<th>AUPRC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Retrospective test set (with layer normalization)</td>
<td>0.816</td>
<td>0.483</td>
</tr>
<tr>
<td>Temporal validation (with layer normalization)</td>
<td>0.862</td>
<td>0.386</td>
</tr>
<tr>
<td>Retrospective test set (without layer normalization)</td>
<td>0.782</td>
<td>0.01</td>
</tr>
</tbody>
</table>
8 Key Decision Points in AI Adoption Process

1. Identify and prioritize a problem
2. Evaluate AI as a viable component of the solution
3. Develop measures of outcomes and success of the AI product
4. Design a new optimal workflow to facilitate integration
5. Evaluate pre-integration safety and effectiveness of the AI product
6. Execute change management, workflow integration, and scaling strategy
7. Monitor and maintain the AI product
8. Update or decommission the AI product
Build Modular Infrastructure to Support Many Projects

Flexible Data Pipeline Technology Infrastructure

**DIHI Data Pipeline**

- **Control & Monitor**
- **Task scheduler**
- **Task Queue**
- **Data Extraction**
- **Data Cleaning**
- **Data Monitoring**

**Other Sources**
- Docker + Kubernetes
- Airflow
- RabbitMQ
- Ansible
- Gitlab

**Secure Environment**
- RWD/E and HSR
- Tableau or Superset
- Operational Dashboards

**Structured data: Raw, cleaned, normalized and standardized**
- DELPHI (clean data)

**Dev/Stage**
- Registry/CDM
- Model Outputs
- Dashboards

**EHR & Other Data sources**
- clarity
- chronicles
- Other Sources

**DIHI Data Pipeline**

- **Data**
  - Data normalization standardization
  - CCS class grouping
  - ICD9 - ICT10 crosswalks
  - Comorbidty groupings
  - Lab grouping
  - Meds therapeutic class
  - Provider grouping

**ML Model-1**

**ML Model-2**

**ML Model-3**

**Operational Dashboards**

**ML Model Outputs**

**Control & Monitor**

**Task scheduler**

**Task Queue**

**Data Extraction**

**Data Cleaning**

**Data Monitoring**

**DIHI Data Pipeline**

- **Docker + Kubernetes**
- **Airflow**
- **RabbitMQ**
- **Ansible**
- **Gitlab**

**Operational Dashboards**

**Model Outputs**

**Dashboards**
Model Labels

**Model Facts**

- **Model name**: Deep Sepsis
- **Locale**: Duke University Hospital

**Approval Date**: 09/22/2019  
**Last Update**: 09/24/2019  
**Version**: 1.0

**Summary**

This model uses EHR input data collected from a patient’s current inpatient encounter to estimate the probability that the patient will meet sepsis criteria within the next 4 hours. It was developed in 2016-2019 by the Duke Institute for Health Innovation. The model was licensed to CareMed in July 2019.

**Mechanism**

- **Outcome**: sepsis within the next 4 hours, see (1) for sepsis criteria
- **Output**: 0% - 100% probability of sepsis occurring in the next 4 hours
- **Patient population**: all adult patients >18 y.o. presenting to DUH ED and admitted
- **Time of prediction**: every hour of a patient’s encounter
- **Input data source**: electronic health record (EHR)
- **Input data type**: demographics, analytes, vitals, medication administrations
- **Training data location and time period**: DUH, 10/2014 – 12/2015
- **Model type**: Recurrent Neural Network

**Validation and performance**

- **Prevalence**
  - Local Retrospective: 18.9%
  - Local Temporal: 6.4%
  - Local Prospective: TBD
- **AUC**
  - Local Retrospective: 0.88
  - Local Temporal: 0.94
  - Local Prospective: TBD
- **PPV @ Sensitivity of 60%**
  - Local Retrospective: 0.14
  - Local Temporal: 0.20
  - Local Prospective: TBD
- **Sensitivity @ PPV of 20%**
  - Local Retrospective: 0.50
  - Local Temporal: 0.66
  - Local Prospective: TBD

**Uses and directions**

- **Operational use case(s)**: Every hour, data is pulled from the EHR to calculate risk of sepsis for every patient at the DUH ED. A rapid response team nurse reviews every high-risk patient with a physician in the ED to confirm whether or not to initiate treatment for sepsis.
- **General use**: This model is intended to be used by clinicians to identify patients for further assessment for sepsis. The model is not a diagnostic for sepsis and is not meant to guide or drive clinical care. This model is intended to complement other pieces of patient information related to sepsis as well as a physical evaluation to determine the need for sepsis treatment.
- **Examples of appropriate decisions to support**: Patient X has a high risk of sepsis according to the model. A rapid response team nurse discusses the patient with the ED physician caring for the patient and they agree the patient does not require treatment for sepsis.
- **Before using this model**: Test the model retrospectively and prospectively on local data to confirm generalizability of the model to the local setting.
- **Safety and efficacy evaluation**: Analysis of data from clinical trial (NCT03655626) underway. Preliminary data shows rapid response team, nurse-driven workflow is effective at improving sepsis treatment bundle compliance.

**Warning**

- **General warnings**: This model was not trained or evaluated on patients receiving care in the ICU. Do not use this model in the ICU setting without further evaluation. This model was trained to identify the first episode of sepsis during an inpatient encounter. During long inpatient stays with multiple sepsis episodes, model accuracy needs to be further evaluated. The model is not interpretable and does not provide rationale for high risk scores. Clinical end users are expected to place model output in context with other clinical information to make final determination of diagnosis.
- **Examples of inappropriate decisions to support**: This model may not be accurate outside of the target population, primarily adults in the non-ICU setting. This model is not a diagnostic and is not designed to guide clinical diagnosis and treatment for sepsis.
- **Discontinue use if**: Clinical staff raise concerns about utility of the model for the indicated use case or large, systematic changes occur at the data level that necessitates re-training of the model.

**Other information**

- **Outcome Definition**: [Link](https://doi.org/10.1101/648907)
- **Related model**: [Link](http://doi.org/10.1001/jama.2016.0288)
- **Model development & validation**: arxiv.org/abs/1708.05894
- **Model implementation**: [Preprint](https://jmir.org/preprint/15182)
- **Clinical trial**: [ClinicalTrials.gov](https://clinicaltrials.gov/ct2/show/NCT03655626)
- **Clinical impact evaluation**: TBD
- **For inquiries and additional information**: Please email mark.sendak@duke.edu
8 Key Decision Points in AI Adoption Process

1. Identify and prioritize a problem
2. Evaluate AI as a viable component of the solution
3. Develop measures of outcomes and success of the AI product
4. Design a new optimal workflow to facilitate integration
5. Evaluate pre-integration safety and effectiveness of the AI product
6. Execute change management, workflow integration, and scaling strategy
7. Monitor and maintain the AI product
8. Update or decommission the AI product
## AI System Monitoring at DIHI

Effective monitoring of AI/ML solutions also requires multidisciplinary combination of technical and human capabilities, including expertise in engineering, data analysis, AI/ML, and clinical domain knowledge employed during the solution development phase.

<table>
<thead>
<tr>
<th>Model Monitoring</th>
<th>Solution Monitoring</th>
<th>Operations Monitoring</th>
</tr>
</thead>
</table>
| • Data quality monitoring  
  • Input data accurate, complete, and up-to-date  
  • Entity/group monitoring  
  • Continuous monitoring  
  • Performance comparison  
  • auroc, auprc wrt. training  
  • Analysis cadence: M/Q/Y  
  • Output drift monitoring  
  • Data distribution  
  • Category distribution | • Outcome monitoring  
  • Project specific measures  
  • Bi-annual for most solutions  
  • Workflow changes  
  • Observation / documentation  
  • Usage monitoring  
  • UI tools/dashboard usage  
  • Secondary data analysis  
  • User feedback  
  • Survey for model & solution usability and refinements | • Alerting & notification  
  • Flexible rules-based engine for alerting  
  • Used in clinical workflow  
  • Email/page/spok/sms etc.  
  • Technical monitoring  
  • Model run times, failures etc.  
  • Service level monitoring  
  • Regulatory & Policy  
  • Compliance monitoring for regulation & Duke policies  
  • Ethical and legal standards |
Solution and Input Data Monitoring

Continuous monitoring to ensure safety and quality of data used in model inputs

**Labs**

**SEP-1 bundle compliance | Sepsis Watch model**

**Medication Administrations**

**Goal-concordant care outcome | HealthGuard model**
8 Key Decision Points in AI Adoption Process

1. Identify and prioritize a problem
2. Evaluate AI as a viable component of the solution
3. Develop measures of outcomes and success of the AI product
4. Design a new optimal workflow to facilitate integration
5. Evaluate pre-integration safety and effectiveness of the AI product
6. Execute change management, workflow integration, and scaling strategy
7. Monitor and maintain the AI product
8. Update or decommission the AI product
# Sepsis Watch Post-Integration Lifecycle Management

<table>
<thead>
<tr>
<th>Monitoring &amp; Evaluation</th>
<th>Update</th>
<th>Operational Management</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Event based</strong></td>
<td>• Debug issues that arise (e.g., data endpoint unexpectedly goes down)</td>
<td>• Customize the UI for different user groups</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Train new versions of the model for new clinical settings</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Update user access</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Update reporting functionalities to support clinician management</td>
</tr>
<tr>
<td><strong>Recurring</strong></td>
<td>• Monitor technical elements of the model and source data in pipeline</td>
<td>• Regularly scheduled maintenance (e.g., update groupers every 6 months)</td>
</tr>
<tr>
<td></td>
<td>• Monitor changes that affects work environment and use of model</td>
<td>• Conduct bi-annual end user training to ensure baseline knowledge of AI system</td>
</tr>
<tr>
<td><strong>Semi-Recurring</strong></td>
<td>• Audit the solution for impact on clinical and operational outcomes and impact on work environment</td>
<td>• Improve the UI (e.g., add comment feature, automatically check boxes)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Scale to different use cases</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Convene governance committee monthly</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Secure ongoing funding for AI system use</td>
</tr>
<tr>
<td><strong>One-off</strong></td>
<td>• Create channels for end users to report issues and provide user support services</td>
<td>• Create process and criteria to scope responses to user requests</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Determine ownership of model (e.g., clinical lead, technical lead)</td>
</tr>
</tbody>
</table>
8 Key Decision Points in AI Adoption Process

1. Identify and prioritize a problem
2. Evaluate AI as a viable component of the solution
3. Develop measures of outcomes and success of the AI product
4. Design a new optimal workflow to facilitate integration
5. Evaluate pre-integration safety and effectiveness of the AI product
6. Execute change management, workflow integration, and scaling strategy
7. Monitor and maintain the AI product
8. Update or decommission the AI product
<table>
<thead>
<tr>
<th>Topic</th>
<th>Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duke Institute for Health Innovation</td>
<td>2 mins</td>
</tr>
<tr>
<td>Health AI Partnership</td>
<td>2 mins</td>
</tr>
<tr>
<td>Safe, Effective, and Equitable AI Translation</td>
<td>20 mins</td>
</tr>
<tr>
<td>Health Equity Across the AI Lifecycle (HEAAL)</td>
<td>8 mins</td>
</tr>
</tbody>
</table>
Health AI Partnership Inaugural Workshop

We are all invited to collaboratively develop a framework that addresses core technology evaluation domains across both case studies. While grounded in two real cases studies, the framework should be generalizable.

The framework should answer the question: “our health system is considering adopting a new solution that uses AI; how do we assess the potential future impact on health inequities?”
### Health AI Partnership Inaugural Workshop

<table>
<thead>
<tr>
<th>Time</th>
<th>Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>case 1: NYP Post-partum depression</strong></td>
<td></td>
</tr>
<tr>
<td>1:10 – 1:20 PM</td>
<td>NYP team presents case 1</td>
</tr>
<tr>
<td>1:20 – 1:50 PM</td>
<td>Breakout group activity  &lt;br&gt;- Participants expect to report back  &lt;br&gt;- Observers can take break or work &lt;br&gt;on activity without need to report back</td>
</tr>
<tr>
<td>1:50 – 2:20 PM</td>
<td>Breakout rooms report back, Q&amp;A</td>
</tr>
<tr>
<td>2:20 – 2:35 PM</td>
<td>Expert panel remarks and discussion</td>
</tr>
<tr>
<td>2:35 – 2:45 PM</td>
<td>NYP team presents case 1 learnings and approach</td>
</tr>
<tr>
<td><strong>case 2: PCCI KnowThyPatient</strong></td>
<td></td>
</tr>
<tr>
<td>3:00 – 3:10 PM</td>
<td>PCCI team presents case 2</td>
</tr>
<tr>
<td>3:10 – 3:40 PM</td>
<td>Breakout group activity  &lt;br&gt;- Participants expect to report back  &lt;br&gt;- Observers can take break or work &lt;br&gt;on activity without need to report back</td>
</tr>
<tr>
<td>3:40 – 4:10 PM</td>
<td>Breakout rooms report back, Q&amp;A</td>
</tr>
<tr>
<td>4:10 – 4:25 PM</td>
<td>Expert panel remarks and Q&amp;A</td>
</tr>
<tr>
<td>4:25 – 4:35 PM</td>
<td>PCCI team presents case 2 learnings and approach</td>
</tr>
</tbody>
</table>
Health System Partners

Duke Health
Hackensack Meridian Health
Jefferson Health
Kaiser Permanente
MAYO CLINIC
NewYork-Presbyterian
OCHIN
PCCI
UCSF
University of Michigan

Interested & Affected Parties

AMA
Berkeley University of California
DLA PIPER
IDEO.ORG
Gordon and Betty Moore Foundation
Patrick J. McGovern Foundation
PLOS

Workshop Feedback

- 77 people attended the workshop (including hosts and the HAIP leadership team), and 30 people provided feedback (~39%)
- Overall Experience: (1 = Not at all, 5=Very much)

<table>
<thead>
<tr>
<th></th>
<th>Satisfaction</th>
<th>Safeness</th>
<th>Contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>4.40</td>
<td>4.63</td>
<td>3.83</td>
</tr>
<tr>
<td>Participant</td>
<td>4.10</td>
<td>4.50</td>
<td>3.70</td>
</tr>
<tr>
<td>Case presenter</td>
<td>4.50</td>
<td>5.00</td>
<td>4.25</td>
</tr>
<tr>
<td>Expert panelist</td>
<td>4.80</td>
<td>4.80</td>
<td>4.00</td>
</tr>
<tr>
<td>Observer</td>
<td>4.45</td>
<td>4.55</td>
<td>3.73</td>
</tr>
</tbody>
</table>
## Framework Development Roles

<table>
<thead>
<tr>
<th>Participant</th>
<th>Role</th>
<th>Responsibilities</th>
</tr>
</thead>
<tbody>
<tr>
<td>C Case study presenters</td>
<td>3 innovation teams that develop and implement AI solutions in healthcare delivery organizations</td>
<td>Curated a case study, presented it at the workshop and tested out the framework</td>
</tr>
<tr>
<td>F Framework developers</td>
<td>Clinician, community representative, computer scientist, project manager, legal expert, and sociotechnical scholar</td>
<td>Created a scaffolding of the framework and contributed to developing its content</td>
</tr>
<tr>
<td>H HAIP leaders</td>
<td>Clinicians, computer scientists, lawyers, and a community organizer</td>
<td>Evaluated the framework and provided feedback</td>
</tr>
<tr>
<td>W Workshop participants</td>
<td>77 stakeholders from 10 healthcare delivery organizations and 4 ecosystem partners with clinical, technical, operational, regulatory, and AI ethics expertise</td>
<td>Contributed to developing the content of the framework</td>
</tr>
<tr>
<td>D Design researchers</td>
<td>Qualitative research scientist, clinical data scientist, and project manager</td>
<td>Facilitated the co-design process by collecting, iterating, and synthesizing data from all participants</td>
</tr>
</tbody>
</table>
Procedures: Co-design

- Divergent thinking process
- Convergent thinking process
Procedures: Co-design
Results: Five assessment domains

- 5 assessment domains evaluated across the span of 8 key decision points of AI adoption process

<table>
<thead>
<tr>
<th>Assessment Domain</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accountability</td>
<td>Ensures that potential adverse impacts of using the AI solution are overseen by specific stakeholders within healthcare delivery organizations who have clear responsibilities.</td>
</tr>
<tr>
<td>Fairness</td>
<td>Ensures that the solution performs equitably across patient subgroups by establishing and evaluating meaningful fairness criteria.</td>
</tr>
<tr>
<td>Fitness for purpose</td>
<td>Ensures that the proposed solution solves the identified problem for patient subgroups.</td>
</tr>
<tr>
<td>Reliability and validity</td>
<td>Ensures that the solution achieves pre-specified performance targets across technical, clinical, and process measures.</td>
</tr>
<tr>
<td>Transparency</td>
<td>Ensures that the processes of model development, implementation, identification of potential risks and harms, and progress towards equity objectives are communicated effectively to end users and patient subgroups.</td>
</tr>
</tbody>
</table>
Results: Procedures

• Detailed step-by-step procedures to conduct in each key decision point

• Procedures tailored to an existing and a new AI solution

<table>
<thead>
<tr>
<th>Key Decision Point</th>
<th># of procedures for an existing AI solution</th>
<th># of procedures for evaluating a newly developed AI solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Identify and prioritize a problem</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2. Define AI product specification</td>
<td>13</td>
<td>5</td>
</tr>
<tr>
<td>3. Develop success measures</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>4. Design AI solution workflow</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>5. Generate evidence of safety, efficacy, and equity</td>
<td>6</td>
<td>11</td>
</tr>
<tr>
<td>6. Execute AI solution rollout</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>7. Monitor the AI solution</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>8. Update or decommission the AI solution</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Total # of procedures</td>
<td>37</td>
<td>34</td>
</tr>
</tbody>
</table>
## Results: Key stakeholders

<table>
<thead>
<tr>
<th>Stakeholder Type</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strategic (S)</td>
<td>Stakeholders who develop strategic plans and make decisions that align with organizational interests</td>
</tr>
<tr>
<td>Operational (O)</td>
<td>Stakeholders who manage workflow and make decisions to integrate</td>
</tr>
<tr>
<td>Clinical (C)</td>
<td>Stakeholders who provide clinical care to patients</td>
</tr>
<tr>
<td>Technical (T)</td>
<td>Stakeholders who develop the model and its infrastructure</td>
</tr>
<tr>
<td>Regulatory (R)</td>
<td>Stakeholders who review the model from regulatory and ethical perspectives</td>
</tr>
<tr>
<td>Patient (P)</td>
<td>Stakeholders who receive clinical care and provide insights on their community experiences</td>
</tr>
<tr>
<td>Clinical champion</td>
<td>Clinical stakeholders who lead the project and provide clinical expertise in model development</td>
</tr>
<tr>
<td>Product manager</td>
<td>Stakeholders who manage the project and communicate with various stakeholders involved in the project</td>
</tr>
</tbody>
</table>
## Results: Data sources

<table>
<thead>
<tr>
<th>Data Source</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Local healthcare retrospective data</td>
<td>Historical healthcare data that is curated within the primary healthcare delivery organization seeking to adopt an AI product. When a model is internally developed, the local healthcare retrospective data set is used for training the model.</td>
</tr>
<tr>
<td>Local healthcare prospective data</td>
<td>Real-time healthcare data that is curated within the primary healthcare delivery organization seeking to adopt an AI product. The local healthcare prospective data set is used for validating a model during a ‘silent trial’ and for using the model in clinical care.</td>
</tr>
<tr>
<td>Local non-healthcare data</td>
<td>Non-healthcare data that is curated within a geographic setting where a healthcare delivery organization is based. The local non-healthcare data can be derived from a variety of external sources, including US Census.</td>
</tr>
<tr>
<td>Training data</td>
<td>Data used for training a model. When the model is externally developed, the training data set contains data from an external source.</td>
</tr>
<tr>
<td>Literature review</td>
<td>Data collected through reviewing previously published scholarly works on a specific topic.</td>
</tr>
<tr>
<td>Qualitative data</td>
<td>Data collected through qualitative research methods, including surveys, focus groups, and interviews.</td>
</tr>
</tbody>
</table>
Health Equity Across the AI Lifecycle (HEAAL) Framework Highlights

If there’s evidence of inequity for the condition of interest in historical data, don’t rely on subgroup performance.
Peripheral Artery Disease Case

PAD in white adults

Positive Case

PAD in black adults

Negative Case
Peripheral Artery Disease Case

PAD in white adults

True prevalence = 4/20 (20%)

PAD in black adults

True prevalence = 4/20 (20%)
Peripheral Artery Disease Case

PAD in white adults

Observed prevalence = 4/20 (20%)

PAD in black adults

Observed prevalence = 2/20 (10%)

Outcomes not observed in patients who face barriers to care
Peripheral Artery Disease Case

PAD in white adults

Observed PPV = 2/8 (25%)
Observed Sensitivity = 2/4 (50%)

Patient predicted at high risk of PAD

PAD in black adults

Observed PPV = 0/8 (0%)
Observed Sensitivity = 0/4 (0%)
Peripheral Artery Disease Case

PAD in white adults

Observed PPV = 2/8 (25%)
Observed Sensitivity = 2/4 (50%)

PAD in black adults

True PPV = 2/8 (25%)
Observed PPV = 0/8 (0%)
Observed Sensitivity = 0/4 (0%)
True Sensitivity = 2/4 (50%)
Peripheral Artery Disease Case

**PAD in white adults**

- Observed PPV = 2/8 (25%)
- Observed Sensitivity = 2/4 (50%)

**PAD in black adults**

- True PPV = 2/8 (25%)
- True Sensitivity = 2/4 (50%)

In this scenario, the model performs worse on Black patients because of a diagnosis inequity. If the diagnosis inequity were addressed, the model performance on Black patients would be the same as on White patients.

In cases like this, you cannot accurately assess model performance within the disadvantaged subgroup. You need to test the model prospectively in a way that addresses inequities to accurately assess performance across advantaged and disadvantaged subgroups.
Engage with our community of practice!

Website healthaipartnership.org
Bibliography


Thank you

mark.sendak@duke.edu
suresh.balu@duke.edu