Artificial Intelligence (AI) – Evaluating the Optimum Role for SDOs – HL7, ISO TC215...

An Inquiry initiated by...
HL7 Electronic Health Record Work Group
Reducing Clinician Burden Project

19 September 2022
Over the past several months, the HL7 EHR WG/Reducing Clinician Burden Project Team received several reports/presentations regarding the substantial potential of Artificial Intelligence (AI) in burden reduction and has begun a dialogue with the HL7 Community in this regard.
- Also recognizing that AI has broad application beyond burden reduction.

This dialogue is open to all interested and materially affected parties.

The EHR WG anticipates that other HL7 WGs (and other SDOs) will join and/or lead this effort going forward.
Link:  https://confluence.hl7.org/pages/viewpage.action?pageId=101353849

Upcoming AI Sessions

AI Evaluation Objectives

AI Presentations

AI Reference Sources – Application in Health/Healthcare

AI Reference Sources – Cross-Industry Application

AI Reference Sources – US Executive Orders w/Followup Activities
Artificial Intelligence

Terms of Note

- Primarily = Artificial Intelligence
- Other terms of note
  - Assistive Intelligence
  - Augmented Intelligence
  - Action-able Intelligence
Identify Key Objectives

- To review pertinent Reference Sources to understand the principles, constructs, constraints and benefits of AI in health (support of individual and population health), healthcare (clinical practice, healthcare administration and operations) and healthcare informatics – all within an overarching framework of trust, integrity, efficacy and safety.

- To assess what role Standards Development Organizations, such as HL7 and ISO TC215, can/should play in development of AI standards.
  - Just Ratified – ISO/HL7 Cooperation Agreement

- To identify key collaborators in the AI evaluation process and any subsequent activities.
AI Model = Algorithm(s) + Data(set(s))

Might we use... HL7 FHIR-based data(set(s)) based on FHIR resources?
AI Engagement in Decision Making and Follow Up

Action Levels (equate to "Comfort Levels")

1) Gather data: real-time
2) Determine applicable algorithm(s): real-time
3) Execute algorithm(s): real-time
4) If “true”, determine applicable action(s): real-time
5) Establish recommended action(s), advise/notify, then:
   a) Stop: Let human decide and initiate action(s); or
   b) Go, unless/until intercepted: Automatically initiate defined action(s), allow human to intervene/cancel; or
   c) Go: Automatically initiate defined action(s)

Initial Action (Comfort) Level: AI Engagement as per Steps 1 thru 5a
“Quality principles include health care that is: safe, effective, patient-centered, timely, efficient, and equitable.

“Healthcare system goals include:

1. “Enhancing patient experience,
2. “Improving population health,
3. “Reducing per capita health care costs, and
4. “Safe-guarding/improving the work-life of health care providers, as well as
5. “Improving business processes and, most recently,
6. “Equity and inclusion.”
“High-level principles, tenants, and abstract requirements for development and deployment of AI... revealed [these] eight themes:

1. “Privacy,
2. “Accountability,
3. “Safety and security,
4. “Transparency and explain-ability,
5. “Fairness and non-discrimination,
6. “Human control of technology,
7. “Professional responsibility, and
8. “Promotion of human values.”
Key Reference Sources

1) Organization of Economic Cooperation and Development - AI Principles, May 2019
2) European Commission - Ethics guidelines for trustworthy AI, 8 Apr 2019
3) US Executive Order 13960 - Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government - US Federal Register, 8 Dec 2020
<table>
<thead>
<tr>
<th>NIST Draft Taxonomy</th>
<th>1) OECD</th>
<th>2) EU</th>
<th>3) US EO 13960</th>
</tr>
</thead>
</table>
| **Technical Design Attributes** | • Accuracy  
• Reliability  
• Robustness  
• Security and Resilience | • Robustness  
• Security | • Technical robustness  
• Accurate, reliable and effective  
• Secure and resilient |
| **Socio-Technical Attributes** | • Explain-ability  
• Interpretability  
• Privacy  
• Safety  
• Absence of Bias | • Safety  
• Privacy  
• Non-discrimination | • Safe  
• Understandable by subject matter experts, users, and others, as appropriate |
| **Guiding Principles Contributing to Trustworthiness** | • Fairness  
• Accountability  
• Transparency | • Traceability to human values  
• Transparency and responsible disclosure  
• Accountability | • Human agency and oversight  
• Data governance  
• Transparency  
• Diversity & fairness  
• Environmental and societal well-being  
• Accountability |
|  |  |  | • Lawful and respectful of our Nations values  
• Responsible and traceable  
• Regularly monitored  
• Transparent  
• Accountable |
“Health-related artificial intelligence needs rigorous evaluation and guardrails”

By John D. Halamka, Suchi Saria and Nigam H. Shah

Published March 17, 2022
John Halamka, Suchi Saria, Nigam Shah

Actionable Suggestions

1. **Create a label for every algorithm** — analogous to a nutrition label, or a drug label — describing the data used to develop an algorithm, its usefulness and limitations, its measured performance, and its suitability for a given population. When you buy a can of soup, you decide if the calories, fat, and sodium align with your needs and preferences. When health systems decide on a drug to use, a medical review board assesses its utility. The same should be true of AI in health care.

2. **Test and monitor the performance** of algorithm-guided care within the settings in which it is deployed in an ongoing way. Testing should include screening for potential demographic-specific losses in accuracy with tools that find error hotspots that can be hidden by average performance metrics.

3. **Create best practices** for establishing the usefulness, reliability, and fairness of AI algorithms that bring together different organizations to develop and test AI on data sets drawn from diverse and representative groups of patients.
4. **Create a standard way** for government, academia, and industry to monitor the behavior of AI algorithms over time.

5. **Understand clinical context and goals** of each algorithm and know what attributes — quality, safety, outcomes, cost, speed, and the like — are being optimized.

6. **Learn how local variations** in lifestyle, physiology, socioeconomic factors, and access to health care affect both the construction and fielding of AI systems and the risk of bias.

7. **Assess the risk** that AI might be used, intentionally or not, to maintain the status quo and reinforce, rather than eliminate, discriminatory policies.

8. **Develop approaches for appropriate clinical use** of AI in combination with human expertise, experience, and judgment, and discourage overreliance on, or unreflective trust of, algorithmic recommendations.
AI algorithms depend on data quality and context
Consider how to bolster accountability, data quality, context and provenance
  - At each key step in the data lifecycle – Capture, share, use
  - FHIR for AI: Provenance Resource Profile?
Each AI Algorithm is established as valid/applicable, e.g., via clinical trials, given/within a particular context

Contexts to be considered:
- Reference context (e.g., ISO 13119)
- Patient context
- Provider context
- Action context
- Problem context
<table>
<thead>
<tr>
<th>Evaluation Stage – Clinical Decision Support – Contexts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Establishment and Application of AI Algorithms</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>AI Algorithm Development, Clinical Trials, Testing</th>
<th>Does it apply?</th>
<th>AI Algorithm Application</th>
</tr>
</thead>
<tbody>
<tr>
<td>Establish Reference Context</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Establish Patient Context</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Establish Provider Context</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Establish Problem Context</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Establish Action Context</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- If Yes  →  Apply/Execute Algorithm in real-time in clinical setting

17

© Health Level Seven and HL7 are registered trademarks of HL7 International, registered with the United States Patent and Trademark Office.
## HL7 AI Evaluation Process

### Initial Project Action Items

<table>
<thead>
<tr>
<th>Project Action Item(s)</th>
<th>Volunteer Leads</th>
</tr>
</thead>
<tbody>
<tr>
<td>Now Affirmed – Cooperation Agreement to Support Collaboration between HL7 and ISO TC215 Task Force 5 on AI</td>
<td>Pat Baird</td>
</tr>
<tr>
<td></td>
<td>Michael Glickman</td>
</tr>
<tr>
<td></td>
<td>Gary Dickinson</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Contexts</th>
<th>Andeas Thurin</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Reference Metadata (using ISO 13119)</td>
<td>Lawrence McKnight MD</td>
</tr>
<tr>
<td>• Patient Context</td>
<td>Lawrence McKnight MD</td>
</tr>
<tr>
<td>• Provider Context</td>
<td>Lincoln Weed JD</td>
</tr>
<tr>
<td>• Problem Context</td>
<td>Lawrence McKnight MD</td>
</tr>
<tr>
<td>• Action Context</td>
<td></td>
</tr>
</tbody>
</table>
Taking the wisdom of our AI experts, we began to compile a list of potential projects.

What is the optimum role for HL7 and other SDOs in this space?

Initial list created…
<table>
<thead>
<tr>
<th>Item</th>
<th>Rank</th>
<th>Candidate Topic/Area of Focus</th>
<th>Key Proponent(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>Contribute SDO expertise at standards development or providing a framework for such. Use SDO consensus process to review and approve AI artifacts/standards.</td>
<td>John Gachago</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>Establish “Contexts” within which AI Algorithms are developed and ultimately applied in health and healthcare.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>a) ISO 13119 – Clinical knowledge resources - Metadata</td>
<td>a) Anders Thurin</td>
</tr>
<tr>
<td></td>
<td></td>
<td>b) Patient/Provider/Action Contexts</td>
<td>b) Lawrence McKnight MD</td>
</tr>
<tr>
<td></td>
<td></td>
<td>c) Problem Context</td>
<td>c) Lincoln Weed JD</td>
</tr>
<tr>
<td>3*</td>
<td>3</td>
<td>Create a label for every algorithm — analogous to a nutrition label, or a drug label — describing the data used to develop an algorithm, its usefulness and limitations, its measured performance, and its suitability for a given population. [Similar to Item 2?]</td>
<td>John Halamka MD, Suchi Saria, Nigam Shah</td>
</tr>
<tr>
<td>4*</td>
<td>4</td>
<td>Test and monitor the performance of algorithm-guided care within the settings in which it is deployed in an ongoing way.</td>
<td>Halamka, Saria, Shah</td>
</tr>
<tr>
<td>5*</td>
<td>5</td>
<td>Create best practices for establishing the usefulness, reliability, and fairness of AI algorithms that bring together different organizations to develop and test AI on data sets drawn from diverse and representative groups of patients.</td>
<td>Halamka, Saria, Shah</td>
</tr>
<tr>
<td>6*</td>
<td>6</td>
<td>Create a standard way for government, academia, and industry to monitor the behavior of AI algorithms over time.</td>
<td>Halamka, Saria, Shah</td>
</tr>
<tr>
<td>7*</td>
<td>7</td>
<td>Understand clinical context and goals of each algorithm and know what attributes — quality, safety, outcomes, cost, speed, and the like — are being optimized.</td>
<td>Halamka, Saria, Shah</td>
</tr>
<tr>
<td>8*</td>
<td>8</td>
<td>Learn how local variations in lifestyle, physiology, socioeconomic factors, and access to health care affect both the construction and fielding of AI systems and the risk of bias.</td>
<td>Halamka, Saria, Shah</td>
</tr>
<tr>
<td>Item</td>
<td>Rank</td>
<td>Candidate Topic/Area of Focus</td>
<td>Key Proponent(s)</td>
</tr>
<tr>
<td>------</td>
<td>------</td>
<td>------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>9</td>
<td>9*</td>
<td>Assess the risk that AI might be used, intentionally or not, to maintain the status quo and reinforce, rather than eliminate, discriminatory policies.</td>
<td>Halamka, Saria, Shah</td>
</tr>
<tr>
<td>10</td>
<td>10*</td>
<td>Develop approaches for appropriate clinical use of AI in combination with human expertise, experience, and judgment, and discourage overreliance on, or unreflective trust of, algorithmic recommendations.</td>
<td>Halamka, Saria, Shah</td>
</tr>
<tr>
<td>11</td>
<td></td>
<td>AI algorithms depend on data quality and context. Consider how to bolster accountability, data quality, context and provenance at each step in the data lifecycle (capture, share, use). Develop Provenance Resource Profile (FHIR for AI)</td>
<td>John Halamka MD Gary Dickinson</td>
</tr>
<tr>
<td>12</td>
<td></td>
<td>Develop standards for auditing AI</td>
<td>John Gachago</td>
</tr>
<tr>
<td>13</td>
<td></td>
<td>Develop standards for CDS software to identify AI tools potentially applicable to a patient problem, combine data about the applicable tools (including but not limited to metadata in ISO 13119) with data about the patient/problem, and generate options (i.e. alternative tools) with details about the pros and cons of each option specific to the patient (see paragraph I.A.4 of the “Problem Context” memo in Item 2c above).</td>
<td>Lincoln Weed JD Chris Weed Charles Burger MD</td>
</tr>
<tr>
<td>14</td>
<td></td>
<td>Guidelines for articles and clinical trial literature, standardized language with markup, optimized for machine learning</td>
<td>Steve Datena MD Jim Sorace MD</td>
</tr>
<tr>
<td>15</td>
<td></td>
<td>Everyday patient care record, rigorously formatted – extracted for diagnosis, treatment, decisions, use for clinical trials</td>
<td>Jim Sorace MD</td>
</tr>
<tr>
<td>16</td>
<td></td>
<td>Data standards for data brokers</td>
<td></td>
</tr>
</tbody>
</table>

Reducing Clinician Burden Project

Schedule

- **Teleconferences**
  - [https://global.gotomeeting.com/join/798931918](https://global.gotomeeting.com/join/798931918)
  - RCB meets 2nd and 4th Mondays each month, 3-4PM ET US
  - POHR meets 1st and 3rd Mondays each month, 3-4PM ET US
  - AI meets 2nd and 4th Tuesdays each month, 4-5PM ET US

  - [https://meet.goto.com/645768157](https://meet.goto.com/645768157)
  - Medication Reconciliation meets alternating Thursdays, 5-6PM ET US

→ Special thanks to Dr Michael Brody for GoToMeeting facilities!
Contact – HL7 EHR WG/Reducing Clinician Burden Project

Co-Facilitators

- Gary Dickinson FHL7: gary.dickinson@ehr-standards.com
- David Schlossman MD PhD FACP MS CPHIMS: dschloss39@gmail.com

- Need AI experts to bolster our leadership team
Contact – HL7 EHR WG
Co-Chairs

- Michael Brody DPM: mbrody@tldsystems.com
  TLD Systems
- Gary Dickinson FHL7: gary.dickinson@ehr-standards.com
  EHR Standards Consulting
- Steve Hufnagel PhD: stephen.hufnagel.hl7@gmail.com
- Mark Janczewski MD: mark.janczewski@gmail.com
  Medical Networks LLC
- John Ritter FHL7: johnritter1@verizon.net
- Michael Van Der Zel: michael.vanderzel@gmail.com
- Pele Yu MD: pele.yu@archildrens.org
  Arkansas Children’s Hospital/University of Arkansas