



Power to Improve Your World

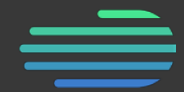
LEVERAGING COLLECTIVE KNOWLEDGE FOR PATIENT IDENTIFICATION AND CARE OPTIMIZATION

National Academy of Medicine DLC
November 30th, 2017



AGENDA

Section
Reviewing Current Processes and Opportunities
Applied Machine Learning, Model Building, and Machine Education
Application for Oncology Patient Identification
Outcomes and Future Implications
Parting Thoughts and Group Questions



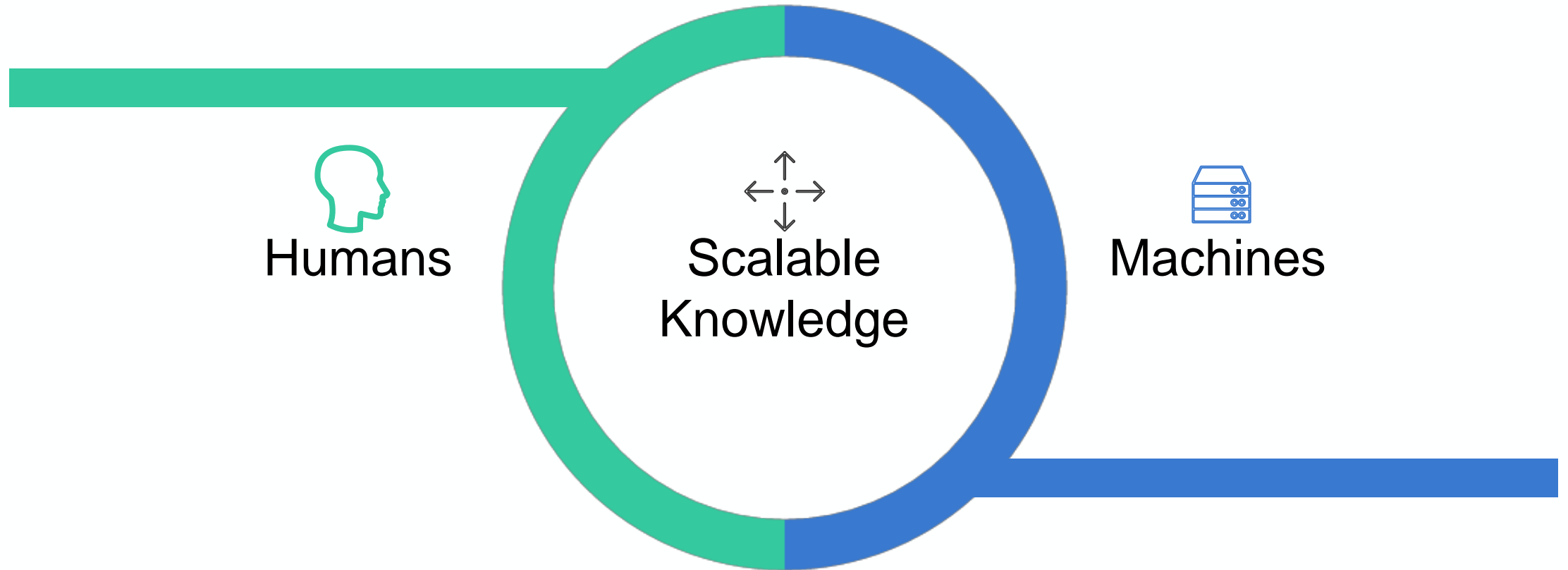
CURRENT PROCESS AND OPPORTUNITIES

Current Care Process and Data Exhaust

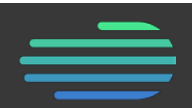
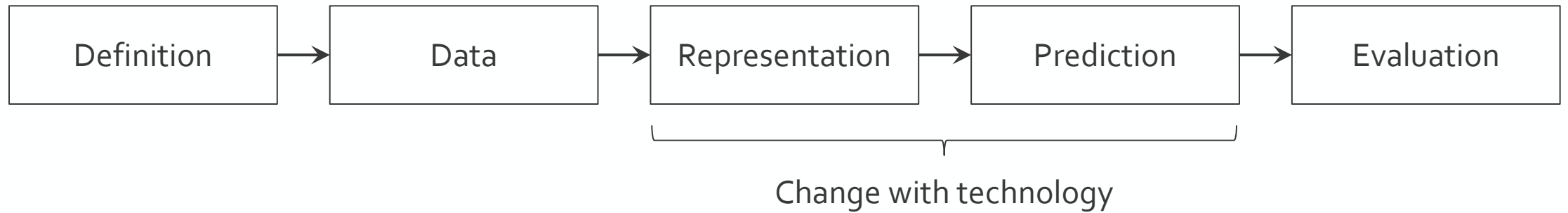
Unstructured Data and the EHR

Data Warehouses and the value of Implicit Knowledge





CORE TENETS OF MODEL BUILDING



DATA



Unlabeled data

Used to learn inherent structure of the problem. Addressed in more detail in “representation”.

Training data

The more data, the better. This is the “curriculum” that educates the machine. Ideal training data consists of the most “useful” data possible. Used to train model parameters.

Validation data

Not used during the training process. Used to help select which “version” of a model is best.

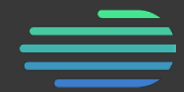
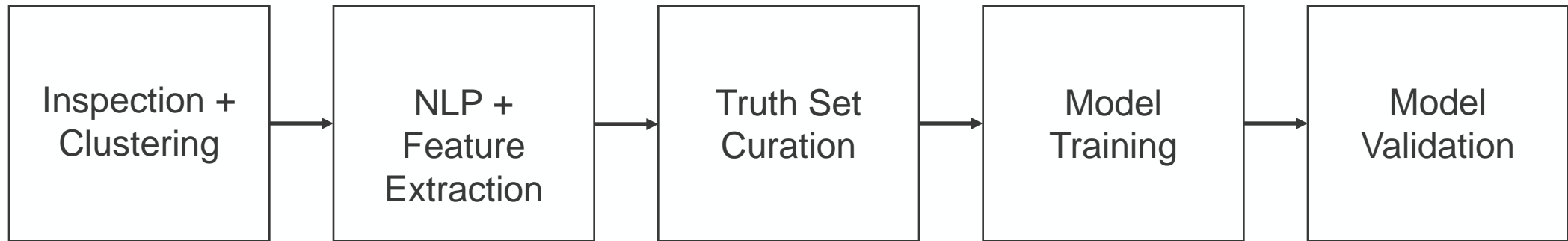
Test data

Helps to test the quality of the final model. Ideal test data is representative of the problem.

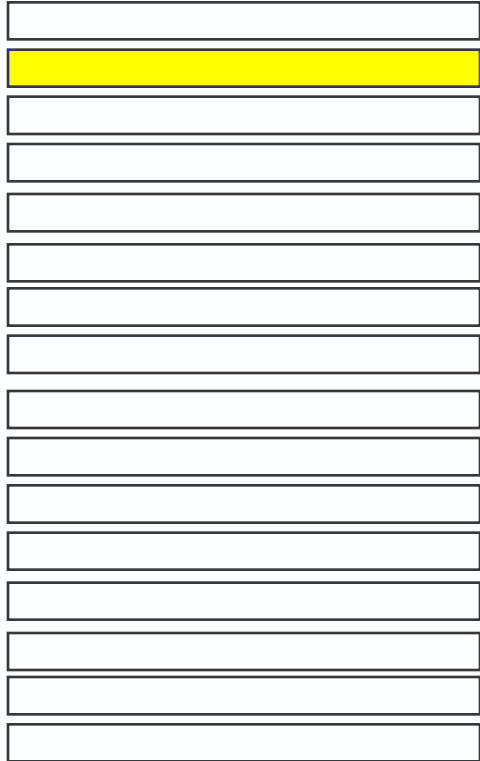
LABELED



DATA SCIENTISTS SOLVE THIS THROUGH MANY TECHNIQUES

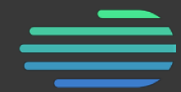


IMBALANCED DATA IS A PROBLEM



If occurrence rate is 0.01% and we need 2,000 positive examples, that requires 20,000,000 examples.

At 10 seconds per example, this is more than 50,000 hours **per model**.



ACTIVE LEARNING – MULTI-CLASS

Annotation Sets > Cancer > Annotate

Cancer Annotations



Search for keywords

CHEST, SINGLE AP PORTABLE SUPINE VIEW: Compared with earlier in the same day, a second Swan-Ganz catheter appears to have been placed. However, there are now multiple tubes and wires overlying the chest and it is very difficult to determine which catheter extends here. At present, one Swan-Ganz catheter appears unchanged, with tip over _____. Another catheter tip overlies a right lower lobe pulmonary artery. The mediastinal contours are prominent, similar to those on the film of 4 hours earlier. A left lower lobe dense consolidation is again seen.

IMPRESSION:

1. Apparent placement of new Swan-Ganz catheter, with tip over right lower pulmonary artery.
2. Dense left lower lobe collapse and/or consolidation. No CHF.

Choose a label:

Breast	1	Stomach	2	Lung	3	Other	4
--------	---	---------	---	------	---	-------	---

Previous

Next

Clear

Training Options

[Train now](#)

[Sampling settings »](#)

Sampler

[Random »](#)

Random samples you haven't yet encountered

Samples Annotated

Total	1005
Breast	301
Stomach	204
Lung	157
Other	343

Performance

	Prec	Rec	F1
Breast	72	95	82
Stomach	68	97	80
Lung	78	95	86

[Cross-validate](#)

ASSISTED LEARNING



Q Cancer

use we are driving
ars that

BC

Training Advice

You might also search for:

cancer

leukemia

chemotherapy

aids

cystic fibrosis

Sample selection method

ONLINE LEARNING

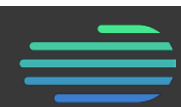
Day-to-day activities feed back into system for next round of training.

The screenshot shows a web application interface with a dark header bar containing navigation links: PROFILES, MONITORING, STUDIO, and INGESTION. The user is logged in as 'admin2'. Below the header, there's a 'Knowledge Graph' dropdown set to 'earnings_reports'. The main content area displays a list of documents under the 'Sent' tab. The list includes columns for 'Sent', 'Senders', and 'Recipients'. Each row shows a document path and a 'WatchList' button. The footer indicates 'Copyright © 2017 Digital Reasoning'.

Sent	Senders	Recipients
04/11/2017 09:22 AM /mnt/david/gen_im/annotated_txt/mnt_gen_im_Transcript_Scraping_POC_JNJ-US-20160518-1842570-C.pdf.xml		
04/11/2017 09:22 AM /mnt/david/gen_im/annotated_txt/mnt_gen_im_Transcript_Scraping_POC_JNJ-US-20150520-1735984-C.pdf.xml		
04/11/2017 09:22 AM /mnt/david/gen_im/annotated_txt/mnt_gen_im_Transcript_Scraping_POC_JNJ-US-20161018-1877833-C.pdf.xml		
04/11/2017 09:22 AM /mnt/david/gen_im/annotated_txt/mnt_gen_im_Transcript_Scraping_POC_JNJ-US-20160419-1821409-C.pdf.xml		
04/11/2017 09:22 AM /mnt/david/gen_im/annotated_txt/mnt_gen_im_Transcript_Scraping_POC_JNJ-US-20150414-1720236-C.pdf.xml		
04/11/2017 09:22 AM /mnt/david/gen_im/annotated_txt/mnt_gen_im_Transcript_Scraping_POC_MRK-DE-20160519-1820346-C.pdf.xml		

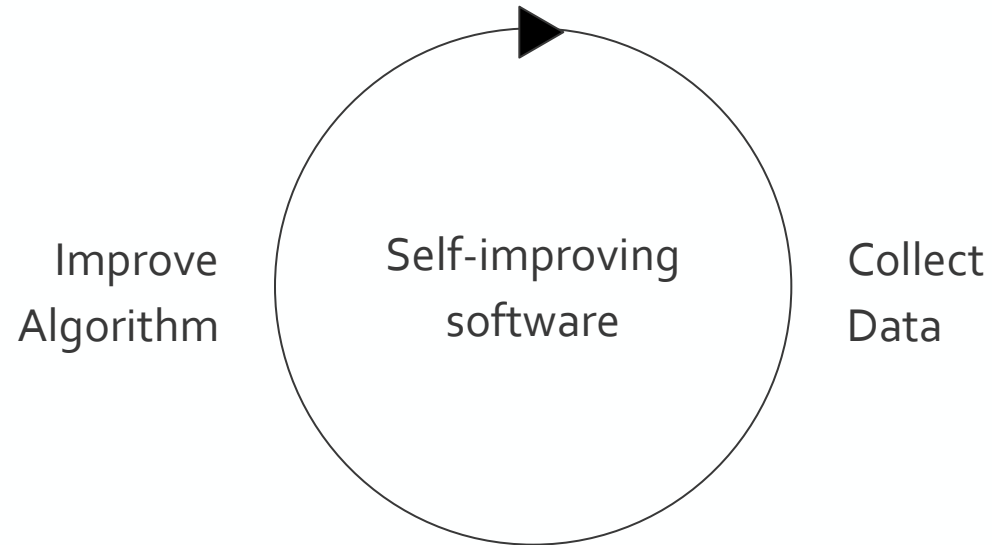
The screenshot shows a web application interface with a dark header bar containing navigation links: PROFILES, MONITORING, STUDIO, and INGESTION. The user is logged in as 'admin2'. Below the header, there's a 'Knowledge Graph' dropdown set to 'earnings_reports'. The main content area displays a table of inflation annotations. The table has columns for 'Label', 'Sample', and 'Annotation Date'. The 'Label' column contains 'Yes' or 'No' values. The 'Sample' column contains text descriptions of inflation. The 'Annotation Date' column contains dates from 2017-05-09. The footer indicates 'Copyright © 2017 Digital Reasoning'.

Label	Sample	Annotation Date
Yes	Price inflation for final goods	2017-05-09
Yes	Overall price inflation remained modest, while wage inflation was moderate.	2017-05-09
No	Overall price inflation was slight, while wage inflation was moderate.	2017-05-09
Yes	Consumer price inflation pressures remained subdued.	2017-05-09
Yes	Prices and Wages Overall price inflation continued to firm somewhat.	2017-05-09
Yes	Wage inflation remained moderate overall.	2017-05-09
Yes	Few signs of inflation were noted.	2017-05-09
Yes	Price inflation remained modest according to contacts.	2017-05-09
Yes	Overall price inflation was modest.	2017-05-09
Yes	Overall price and wage inflation remained modest.	2017-05-09
Yes	Retail price inflation edged up in Richmond.	2017-05-09
Yes	Overall price and wage inflation remained modest.	2017-05-09
Yes	Most respondents expect inflation to remain stable, although roughly one-third anticipate an increase in inflation in coming quarters.	2017-05-09
Yes	Rising gas prices and inflation remain a concern.	2017-05-09

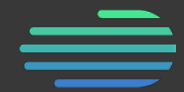


VIRTUOUS CYCLE IMPROVES MODEL OVER TIME

By capturing expert interactions, we continually move algorithms towards human parity over time.



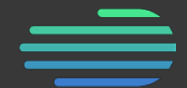
Key enablers: Continuous Learning, Deep Learning (CNNs, LSTMs, Attention Models)



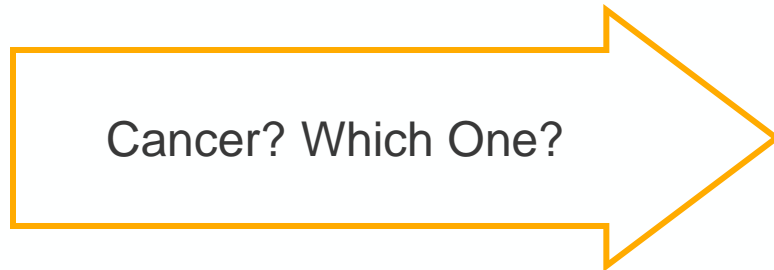


ONCOLOGY PATIENT IDENTIFICATION

Cognitive Computing for better patient and economic outcomes



COGNITIVE-ENABLED CANCER DETECTION FROM UNSTRUCTURED DATA



PATIENTS



PATHOLOGY REPORTS

CANCER REPORTING

Care centers are required by federal law to report cancer patient data to the CDC

CARE COORDINATION

Assisting a patient through their cancer journey, including coordinating visits, communicating, and support



**SOLVES BACKLOG
OF REPORTS**



**REDUCES THE # OF
REGISTRARS**



**REDUCES TIME TO
CONTACT PATIENTS**



**SPEED TREATMENT &
PREVENT CHURN**

CANCER PATIENT ID OUTCOMES

The model correctly identifies:

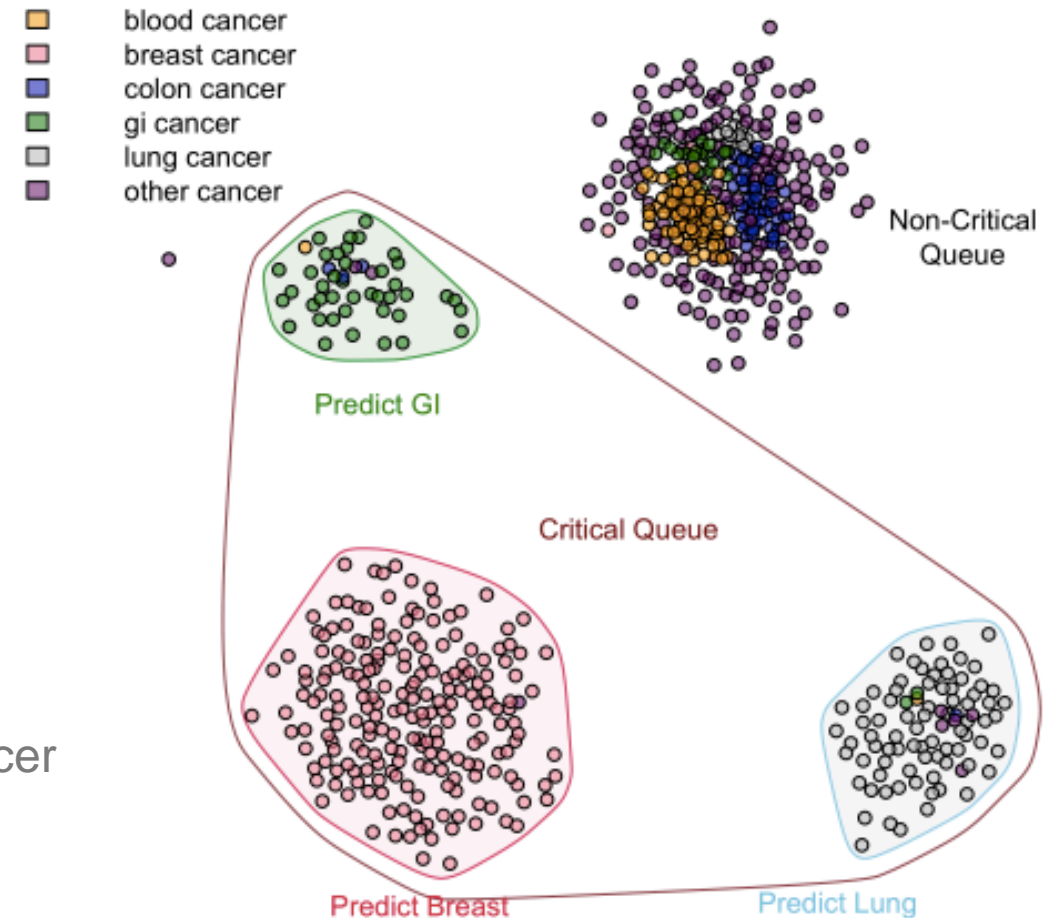
- 90.81%** of positive reports.
- 97.52%** of positive patients.
- 96.15%** of positive patients within 48 hours of diagnosis (with reports in that period).
- 91.53%** of cancer types in positive reports

Outcomes of Production Use:

Efficiency Gains: 2347 hours of reviews reduced to 106 hours (**39X improvement**)

Approximately Equivalent productivity to hiring **118** Cancer Registrars or Navigators on ~20

78k Pathology reports review; **5k+ patients**; now running in **150 hospital systems**



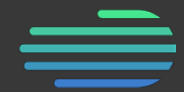
FUTURE IMPLICATIONS

Large Portfolio of Models

Running against many Information Flows

Creating holistic and persistent patient identification and care optimization

That leverages the Collective Wisdom of our Medical Professionals



PARTING THOUGHTS

Machine Education over Data Science – how do we accelerate knowledge transfer?

Open vs. Closed Systems – how do we account for differences across groups and populations

Privacy and Sharing – how do we safely learn across all systems?

