# Human-AI Teaming in Healthcare Data Analytics

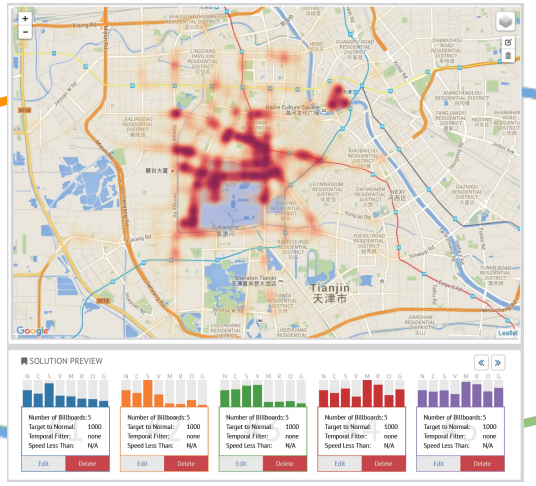
Dongyu Liu

Assistant Professor
Department of Computer Science
University of California, Davis
dyuliu@ucdavis.edu

UCDAVIS
COMPUTER SCIENCE

# My Research

AI System
(Human-Centered)




Visual
Interface

Human

# Your experience in

- **AI (ML)**
- **data visualization**

# AI (ML)

The term is coined by Prof. John McCarthy in 1950s

”The science and engineering of making intelligent machines”

https://hai.stanford.edu/sites/default/files/2020-09/AI-Definitions-HAI.pdf

**Intelligence** might be defined as the ability to learn and perform suitable techniques to solve problems and achieve goals, appropriate to the context in an uncertain, ever-varying world. A fully pre-programmed factory robot is flexible, accurate, and consistent but not intelligent.

**Artificial Intelligence (AI)**, a term coined by emeritus Stanford Professor John McCarthy in 1955, was defined by him as “the science and engineering of making intelligent machines”. Much research has humans program machines to behave in a clever way, like playing chess, but, today, we emphasize machines that can learn, at least somewhat like human beings do.

**Autonomous systems** can independently plan and decide sequences of steps to achieve a specified goal without micro-management. A hospital delivery robot must autonomously navigate busy corridors to succeed in its task. In AI, autonomy doesn’t have the sense of being self-governing common in politics or biology.

**Machine Learning (ML)** is the part of AI studying how computer agents can improve their perception, knowledge, thinking, or actions based on experience or data. For this, ML draws from computer science, statistics, psychology, neuroscience, economics and control theory.

In **supervised learning**, a computer learns to predict human-given labels, such as dog breed based on labeled dog pictures; **unsupervised learning** does not require labels, sometimes making its own prediction tasks such as trying to predict each successive word in a sentence; **reinforcement learning** lets an agent learn action sequences that optimize its total rewards, such as winning games, without explicit examples of good techniques, enabling autonomy.

**Deep Learning** is the use of large multi-layer **(artificial) neural networks** that compute with continuous (real number) representations, a little like the hierarchically organized neurons in human brains. It is currently the most successful ML approach, usable for all types of ML, with better generalization from small data and better scaling to big data and compute budgets.

An **algorithm** lists the precise steps to take, such as a person writes in a computer program. AI systems contain algorithms, but often just for a few parts like a learning or reward calculation method. Much of their behavior emerges via learning from data or experience, a sea change in system design that Stanford alumnus Andrej Karpathy dubbed **Software 2.0**.

**Narrow AI** is intelligent systems for one particular thing, e.g., **speech** or **facial recognition**. **Human-level AI**, or **Artificial General Intelligence (AGI)**, seeks broadly intelligent, context-aware machines. It is needed for effective **social chatbots** or **human-robot interaction**.

**Human-Centered Artificial Intelligence** is AI that seeks to augment the abilities of, address the societal needs of, and draw inspiration from human beings. It researches and builds effective partners and tools for people, such as a robot helper and companion for the elderly.

*Text by Professor Christopher Manning, September 2020*

# AI (ML)

The term is coined by Prof. John McCarthy in 1950s

”The science and engineering of making intelligent machines”

https://hai.stanford.edu/sites/default/files/2020-09/AI-Definitions-HAI.pdf

**Intelligence** might be defined as the ability to learn and perform suitable techniques to solve problems and achieve goals, appropriate to the context in an uncertain, ever-varying world. A fully pre-programmed factory robot is flexible, accurate, and consistent but not intelligent.

**Artificial Intelligence (AI)**, a term coined by emeritus Stanford Professor John McCarthy in 1955, was defined by him as “the science and engineering of making intelligent machines”. Much research has humans program machines to behave in a clever way, like playing chess, but, today, we emphasize machines that can learn, at least somewhat like human beings do.

**Autonomous systems** can independently plan and decide sequences of steps to achieve a specified goal without micro-management. A hospital delivery robot must autonomously navigate busy corridors to succeed in its task. In AI, autonomy doesn’t have the sense of being self-governing common in politics or biology.

**Machine Learning (ML)** is the part of AI studying how computer agents can improve their perception, knowledge, thinking, or actions based on experience or data. For this, ML draws from computer science, statistics, psychology, neuroscience, economics and control theory.

In **supervised learning**, a computer learns to predict human-given labels, such as dog breed based on labeled dog pictures; **unsupervised learning** does not require labels, sometimes making its own prediction tasks such as trying to predict each successive word in a sentence; **reinforcement learning** lets an agent learn action sequences that optimize its total rewards, such as winning games, without explicit examples of good techniques, enabling autonomy.

**Deep Learning** is the use of large multi-layer **(artificial) neural networks** that compute with continuous (real number) representations, a little like the hierarchically organized neurons in human brains. It is currently the most successful ML approach, usable for all types of ML, with better generalization from small data and better scaling to big data and compute budgets.

An **algorithm** lists the precise steps to take, such as a person writes in a computer program. AI systems contain algorithms, but often just for a few parts like a learning or reward calculation method. Much of their behavior emerges via learning from data or experience, a sea change in system design that Stanford alumnus Andrej Karpathy dubbed **Software 2.0**.

**Narrow AI** is intelligent systems for one particular thing, e.g., **speech** or **facial recognition**. **Human-level AI**, or **Artificial General Intelligence (AGI)**, seeks broadly intelligent, context-aware machines. It is needed for effective **social chatbots** or **human-robot interaction**.

**Human-Centered Artificial Intelligence** is AI that seeks to augment the abilities of, address the societal needs of, and draw inspiration from human beings. It researches and builds effective partners and tools for people, such as a robot helper and companion for the elderly.

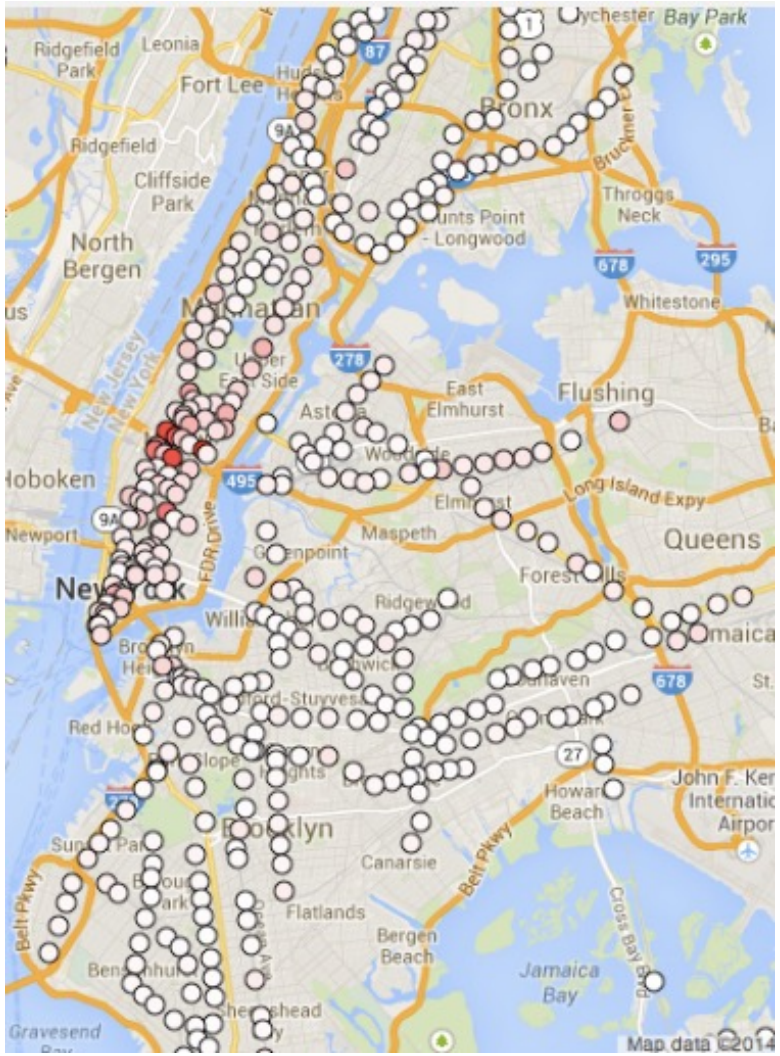
*Text by Professor Christopher Manning, September 2020*

# Data visualization

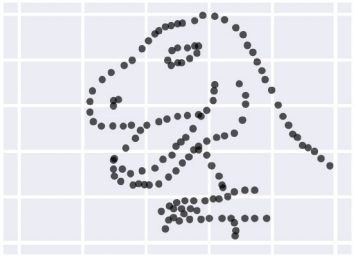
- “Data Visualization is the creation and study of the visual representation of data” - wiki

| | |
|---|---|
| Input | data |
| Output | visual form |
| Goal | insight |

MTA FARE @ NY City

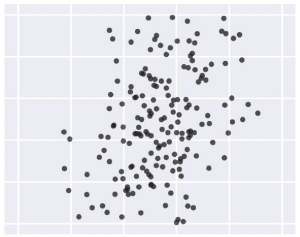
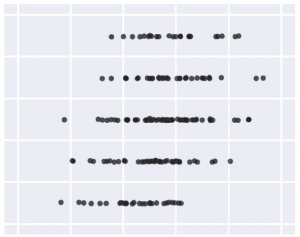
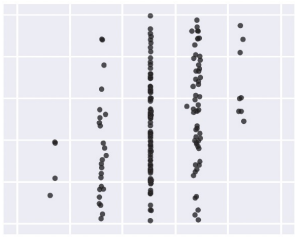
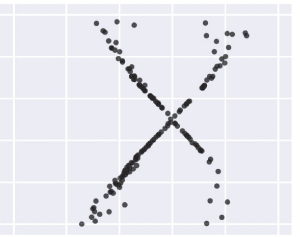
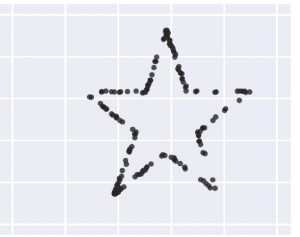
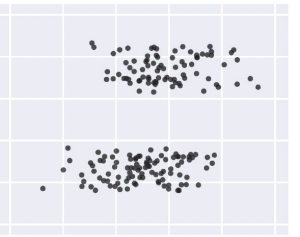
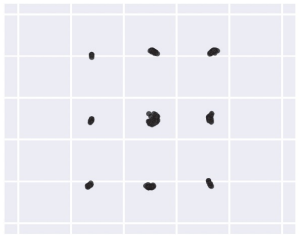
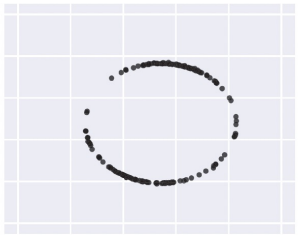
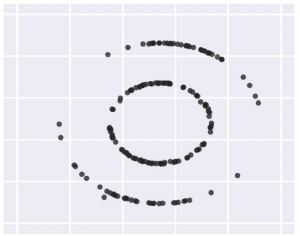
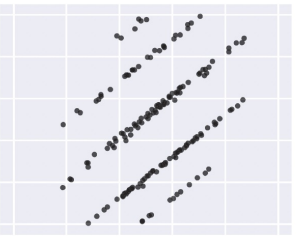
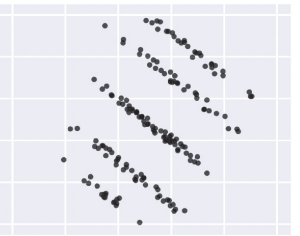
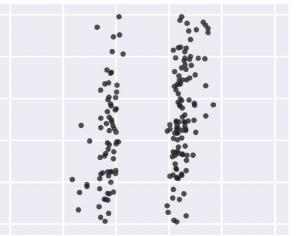


# Same stats, different graphs



X Mean: 54.26 X SD: 16.76

Y Mean: 47.83 Y SD: 26.93 Corr. : -0.06

Matejka, and Fitzmaurice. Same stats, different graphs: generating datasets with varied appearance and identical statistics through simulated annealing. CHI 2017..

# Intended Learning Outcomes (ILOs)

By the end of this class, you will be able to

1. **Describe** what is the data science life cycle
2. **Recognize** the importance of visualization-empowered human AI teaming in such a life cycle
3. **Explain** three key scenarios how human AI teaming can help with healthcare data analytics

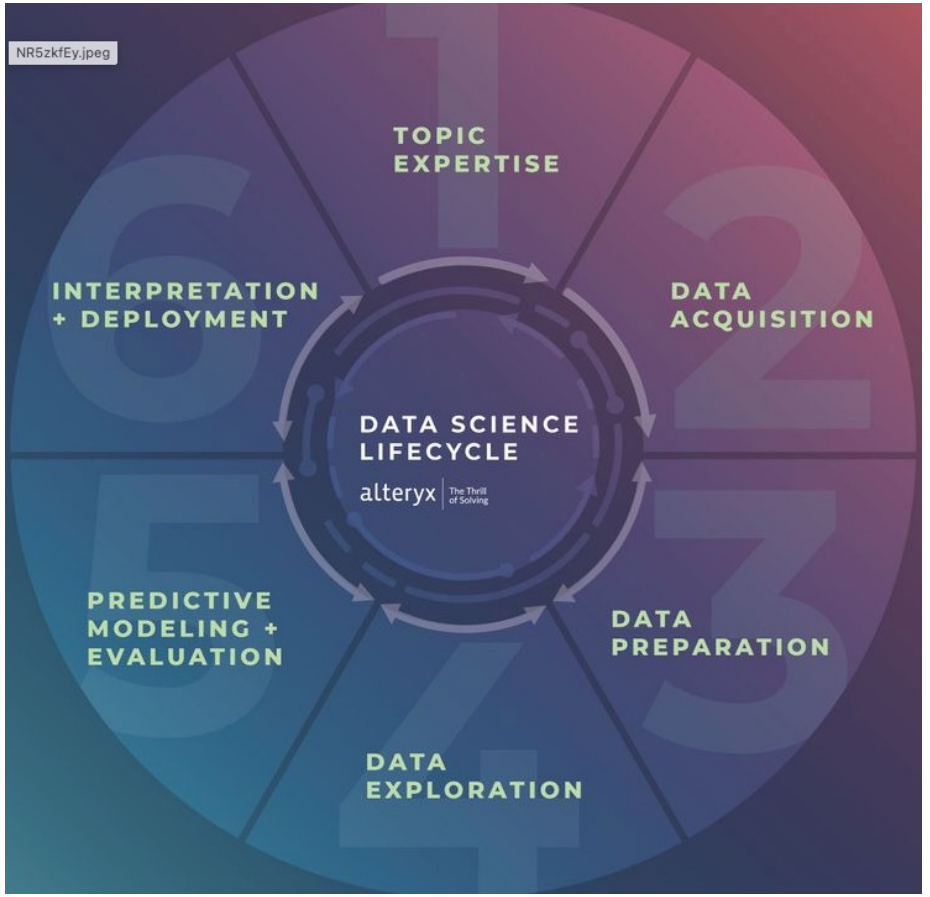

https://karriere.fresenius.de/en-US/digital-careers/focus-on-data-science

# What is a data science life cycle?

**Think-Pair-Share (2 min)**

# Data science life cycle

https://community.alteryx.com/t5/Data-Science/The-Data-Science-Lifecycle/ba-p/408625

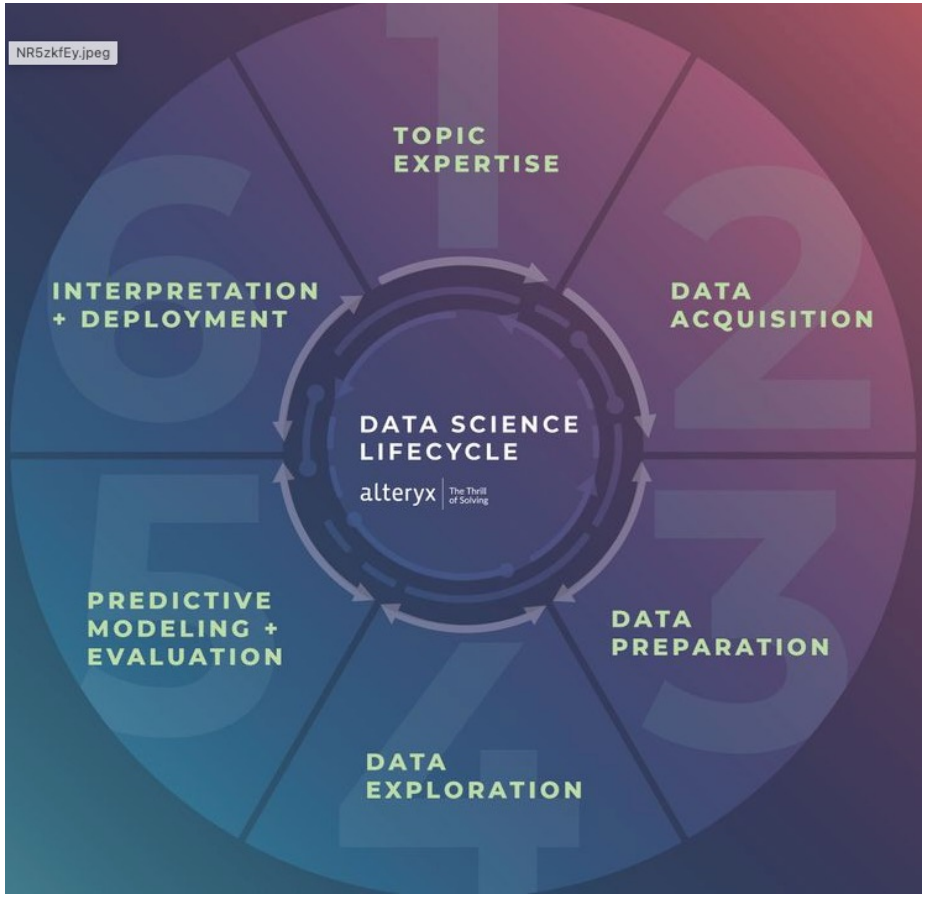



1 The answers from ChatGPT:

- **Problem Definition**: Understand and define the business problem or question you're trying to address.
- **Goal Setting**: Determine the goals and objectives of the project.

# Data science life cycle

https://community.alteryx.com/t5/Data-Science/The-Data-Science-Lifecycle/ba-p/408625

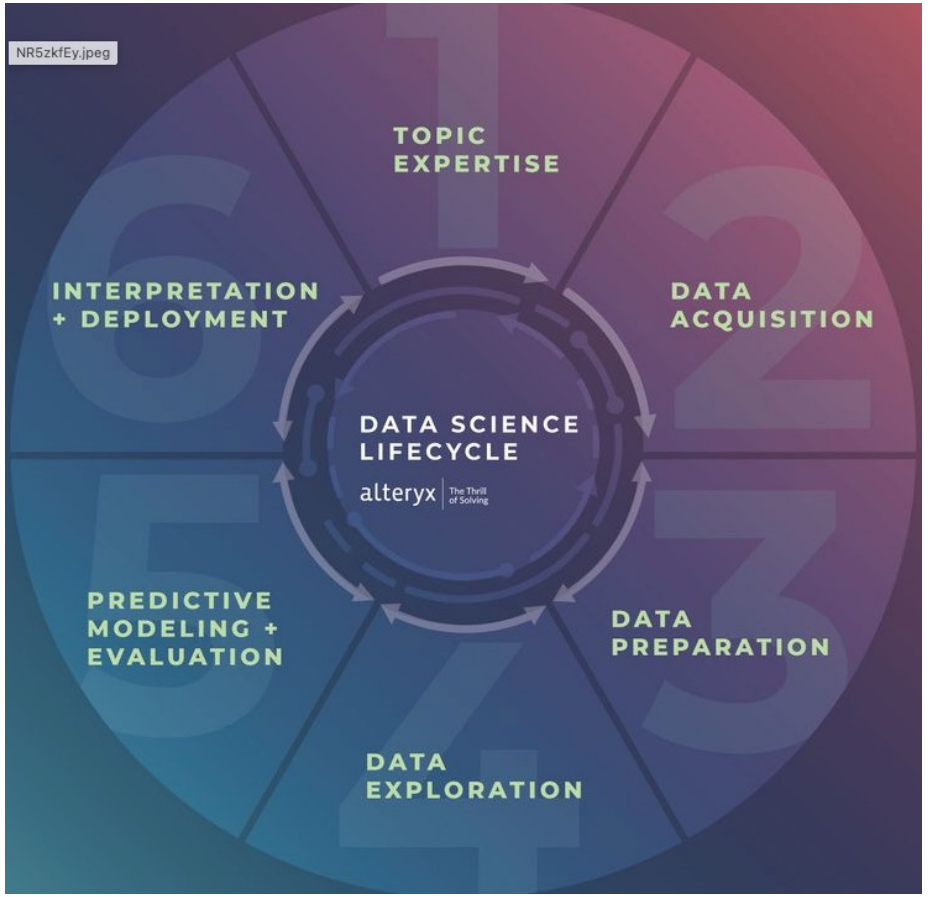



2 The answers from ChatGPT:

- **Data Collection**: Gather data from various sources such as databases, APIs, files, external data sources, etc.

# Data science life cycle

https://community.alteryx.com/t5/Data-Science/The-Data-Science-Lifecycle/ba-p/408625

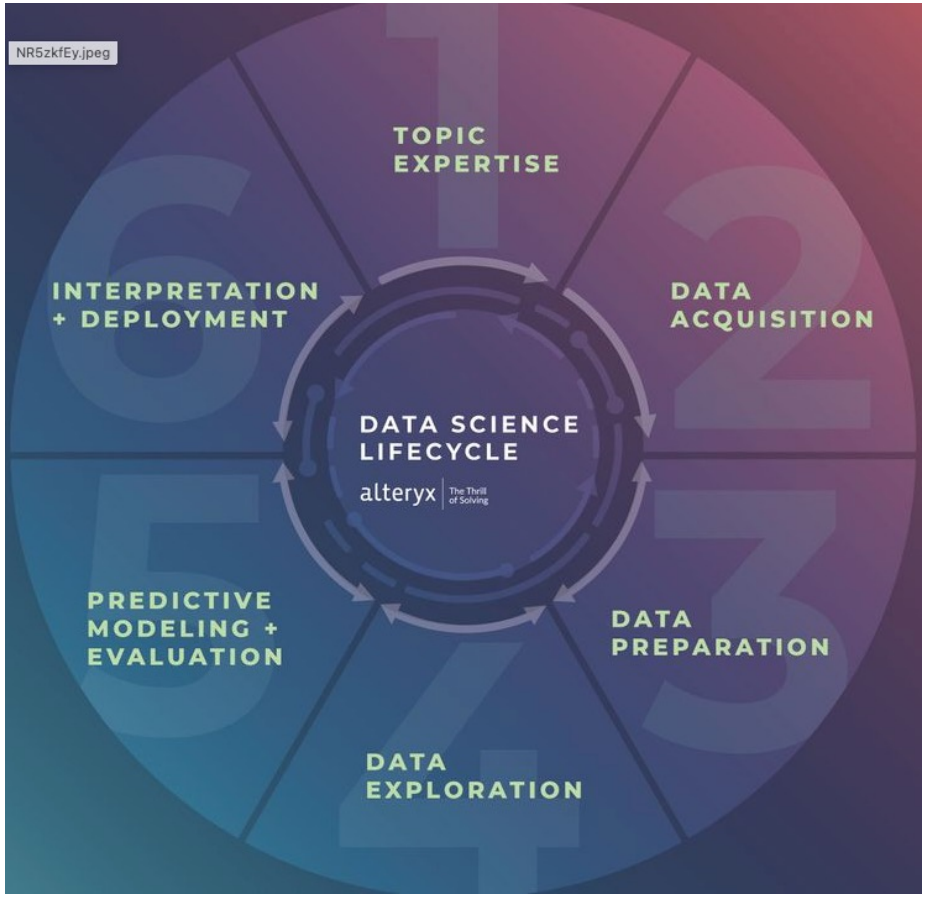



3 The answers from ChatGPT:

- **Data Cleaning**: Deal with missing values, outliers, and any incorrect data.
- **Data Transformation**: Convert the data into a format suitable for modeling, which might include normalizing, scaling, encoding, and feature engineering.

# Data science life cycle

https://community.alteryx.com/t5/Data-Science/The-Data-Science-Lifecycle/ba-p/408625

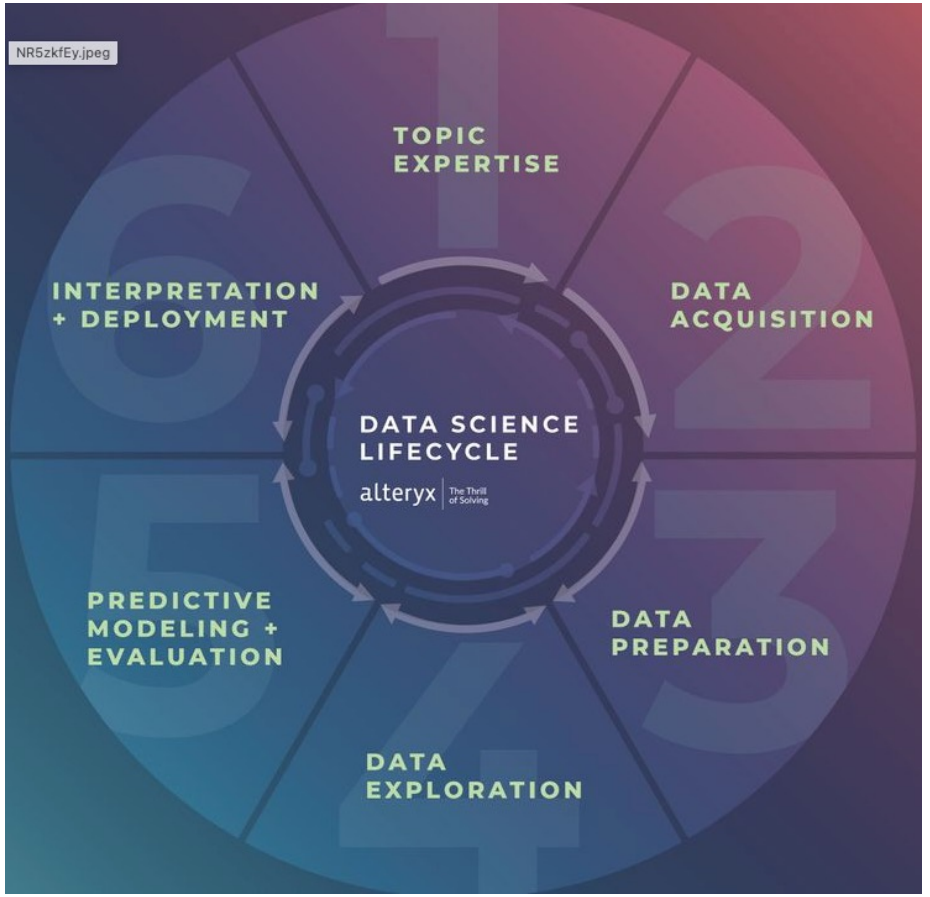



4 The answers from ChatGPT:

- **Data Exploration**: Examine the data to get a sense of its structure, quality, and potential value.
- **Data Quality Assessment**: Identify and address missing data, outliers, and errors.

# Data science life cycle

https://community.alteryx.com/t5/Data-Science/The-Data-Science-Lifecycle/ba-p/408625



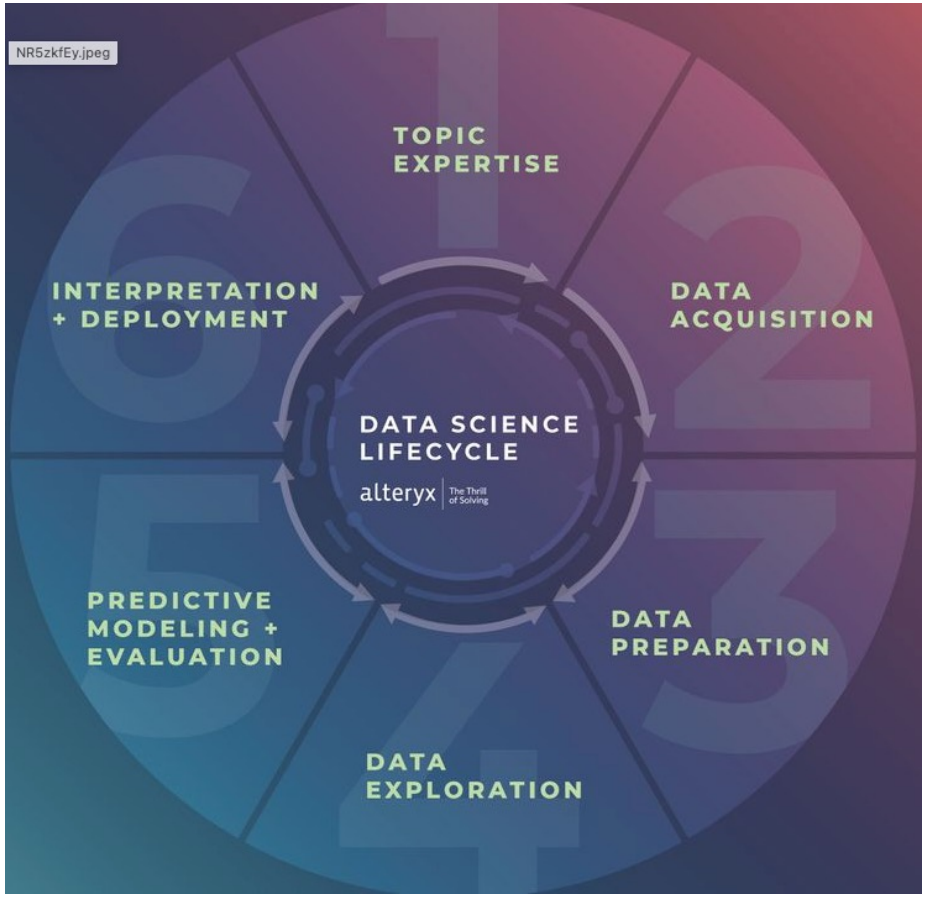

5 The answers from ChatGPT:

- **Feature Selection**: Determine which features (or variables) in the data will be used for modeling.
- **Model Selection**: Choose the appropriate algorithms or models to apply based on the problem type (e.g., regression, classification).
- **Training**: Use a subset of the data to "teach" or "train" the model.
- **Validation**: Evaluate the model's performance on a separate subset of data that it hasn't seen during training.

- **Model Testing**: Test the model's performance on a new dataset.
- **Performance Metrics**: Use relevant metrics (e.g., accuracy, precision, recall, RMSE) to assess the model's quality and reliability.

# Data science life cycle

https://community.alteryx.com/t5/Data-Science/The-Data-Science-Lifecycle/ba-p/408625

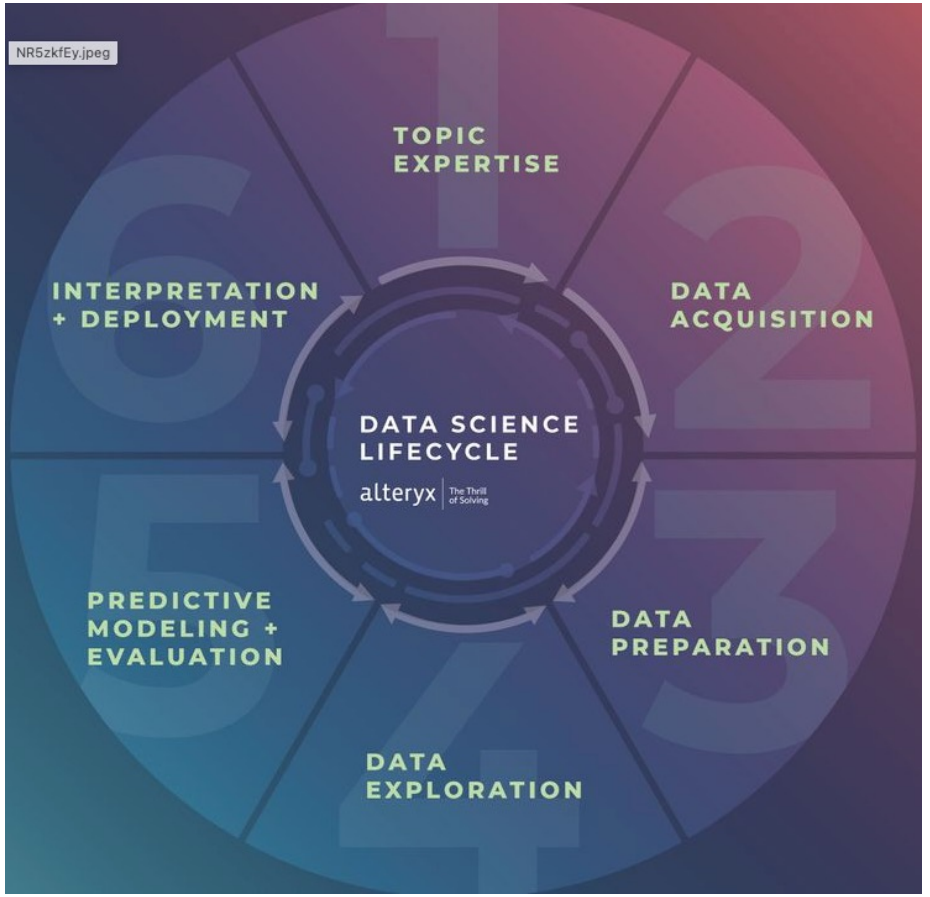



6 The answers from ChatGPT:

- **Deployment Planning**: Determine how the model will be integrated into the existing business processes or systems.
- **Model Deployment**: Implement the model in the desired production environment.
- **User Acceptance**: Ensure stakeholders are comfortable and satisfied with the deployed model's predictions or insights.

# Data science life cycle

https://community.alteryx.com/t5/Data-Science/The-Data-Science-Lifecycle/ba-p/408625




**Iterate**:

- After deployment, new questions, challenges, or insights can emerge, leading to refinements or entirely new cycles of analysis.


AFOSR-3223

*Summary Report*

**AUGMENTING HUMAN INTELLECT: A CONCEPTUAL FRAMEWORK**

*Prepared for:*

DIRECTOR OF INFORMATION SCIENCES
AIR FORCE OFFICE OF SCIENTIFIC RESEARCH
WASHINGTON 25, D.C.

CONTRACT AF 49(638)-1024

*By: D. C. Engelbart*

STANFORD RESEARCH INSTITUTE

MENLO PARK, CALIFORNIA

*SRI

J.C.R. Licklider
March 1960

Douglas C. Engelbart
October 1962

# Augmenting human intelligence with AI

- Perception
- Attention
- Memory
- Language
- Reasoning
- Problem-solving
- Decision-making
- Creativity


Lung cancer detection

Schultheiss, Manuel, et al. "A robust convolutional neural network for lung nodule detection in the presence of foreign bodies." Scientific Reports 10.1 (2020): 12987.

# Augmenting human intelligence with AI

- Perception
- Attention
- Memory
- Language
- Reasoning
- Problem-solving
- Decision-making
- Creativity

### Example of USMLE-style question

A 32-year-old woman comes to the physician because of fatigue, breast tenderness, increased urinary frequency, and intermittent nausea for 2 weeks. Her last menstrual period was 7 weeks ago. She has a history of a seizure disorder treated with carbamazepine. Physical examination shows no abnormalities. A urine pregnancy test is positive. The child is at greatest risk of developing which of the following complications?

A. Renal dysplasia

B. Meningocele

C. Sensorineural hearing loss

D. Vaginal clear cell carcinoma

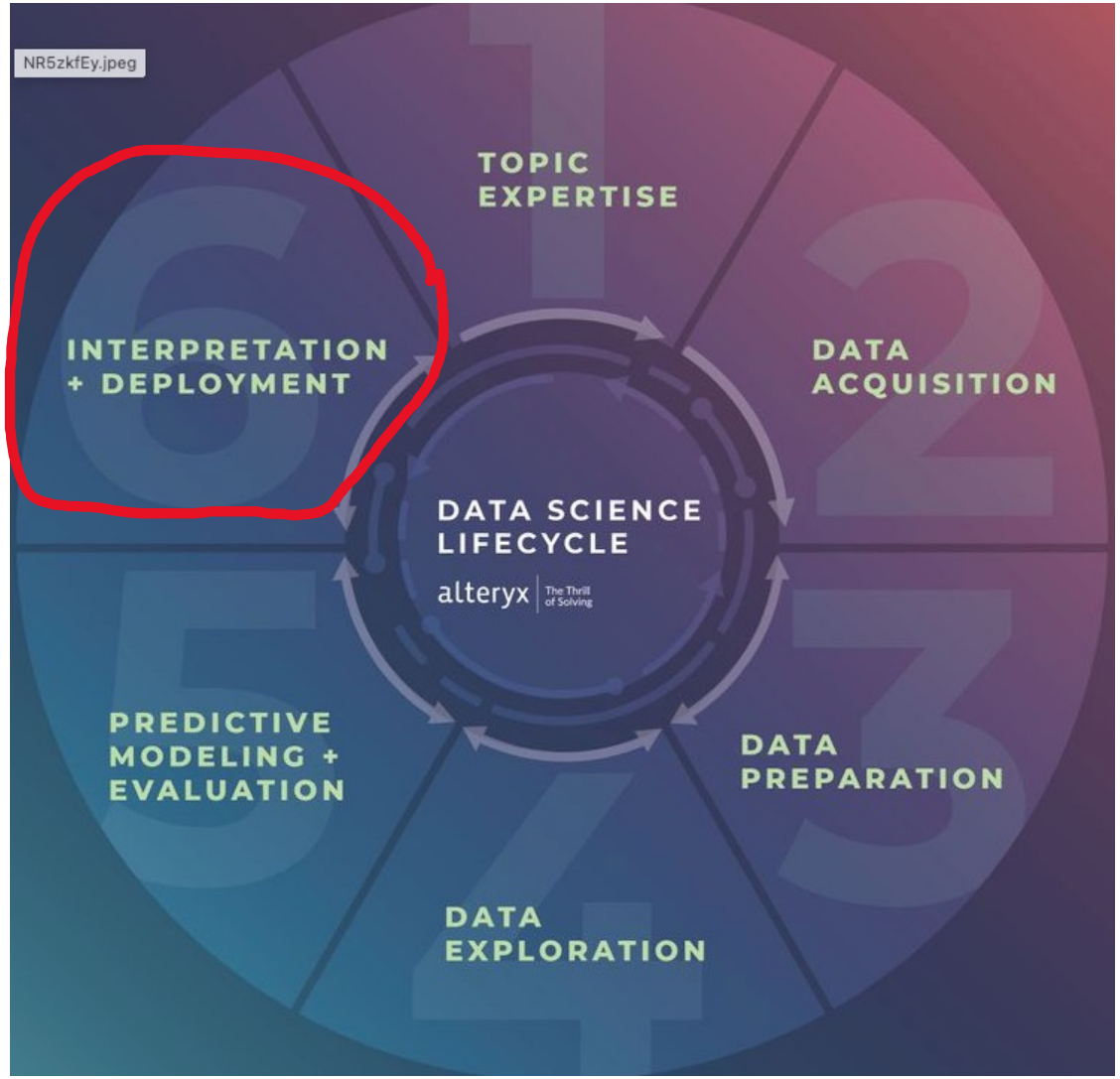
Large language model for medical question answering

Med-PaLM from Google Research. https://sites.research.google/med-palm/

# Augmenting human intelligence with AI

- Perception
- Attention
- Memory
- Language
- Reasoning
- Problem-solving
- Decision-making
- Creativity

Clinical decision support

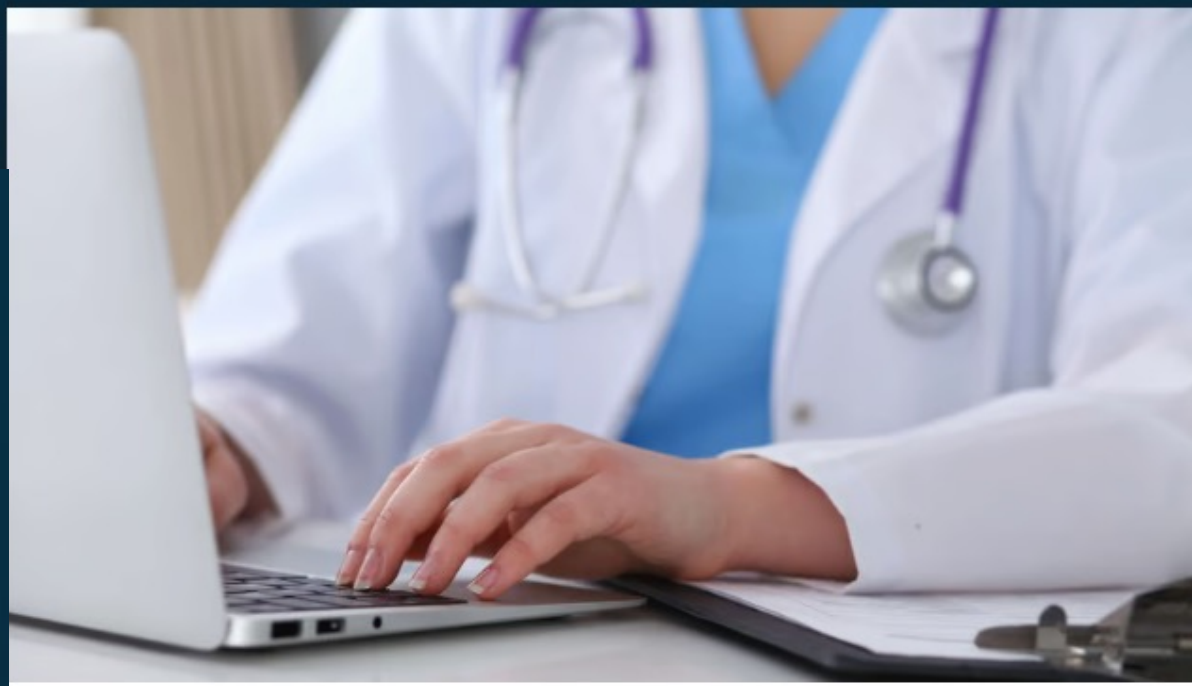

NR5zkfEy.jpeg
TOPIC EXPERTISE
DATA ACQUISITION
DATA PREPARATION
DATA EXPLORATION
PREDICTIVE MODELING + EVALUATION
INTERPRETATION + DEPLOYMENT
DATA SCIENCE LIFECYCLE
alteryx
The Thrill of Solving

# AI is promising, but …

## Nearly half of U.S. doctors say they are anxious about using AI-powered software: survey

By **Heather Landi** • Apr 25, 2019 10:55am

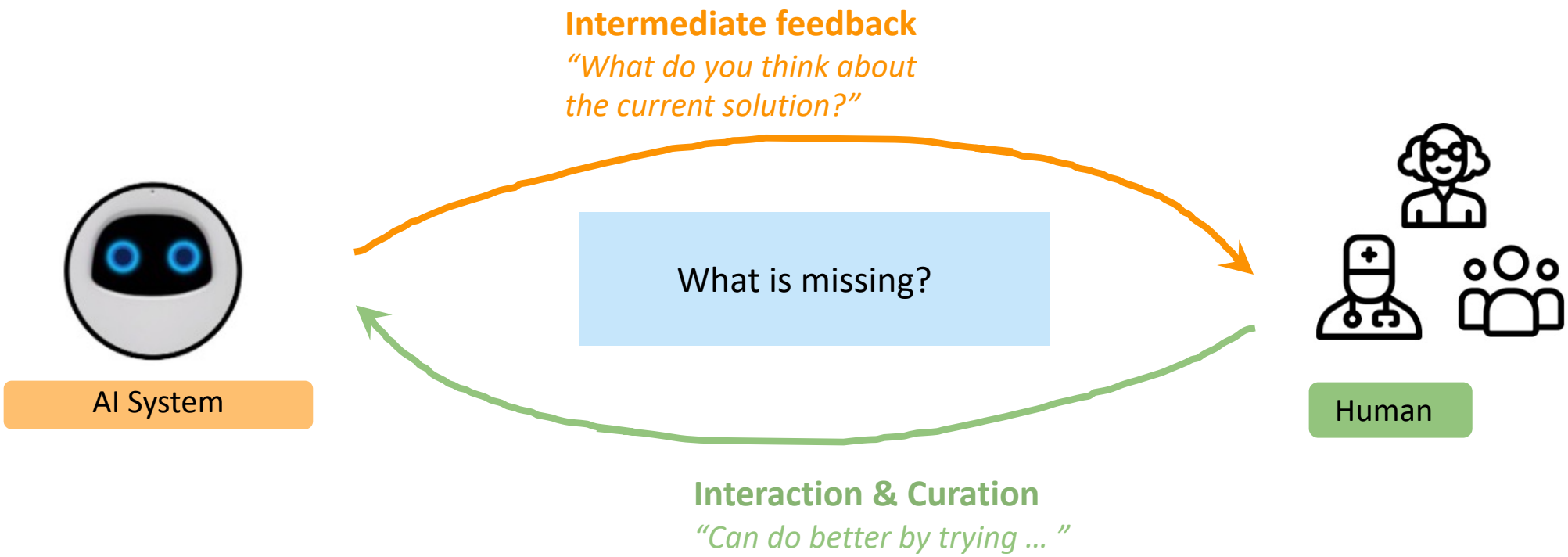
Artificial Intelligence | Clinical Decision Support | Digital health | machine learning



A new physician survey indicates artificial intelligence applications are still in their infancy and have not affected mainstream physician practice at scale. (Getty/andrei_r)

https://www.fiercehealthcare.com/practices/nearly-half-u-s-doctors-say-they-are-anxious-about-using-ai-powered-software-survey

# A general blueprint of Human-AI teaming



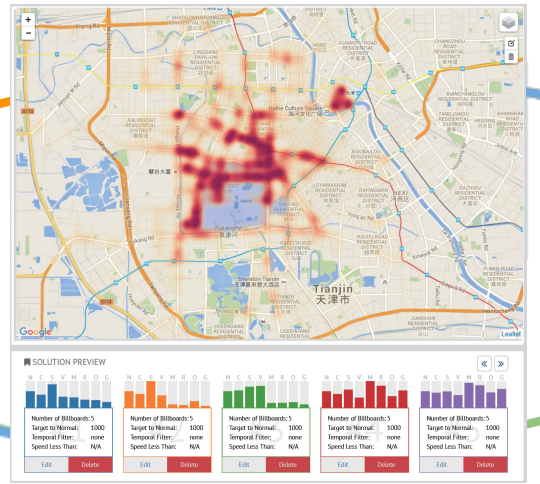

General blueprint for a human-in-the-loop interactive AI system. Image modified from: https://hai.stanford.edu/news/humans-loop-design-interactive-ai-systems

# Visualization-powered teaming workflow

**Intermediate feedback**

*"What do you think about the current solution"*

AI System (Human-Centered)




Visual Interface

Human

**Interaction & Curation**

*"Can do better by trying … "*

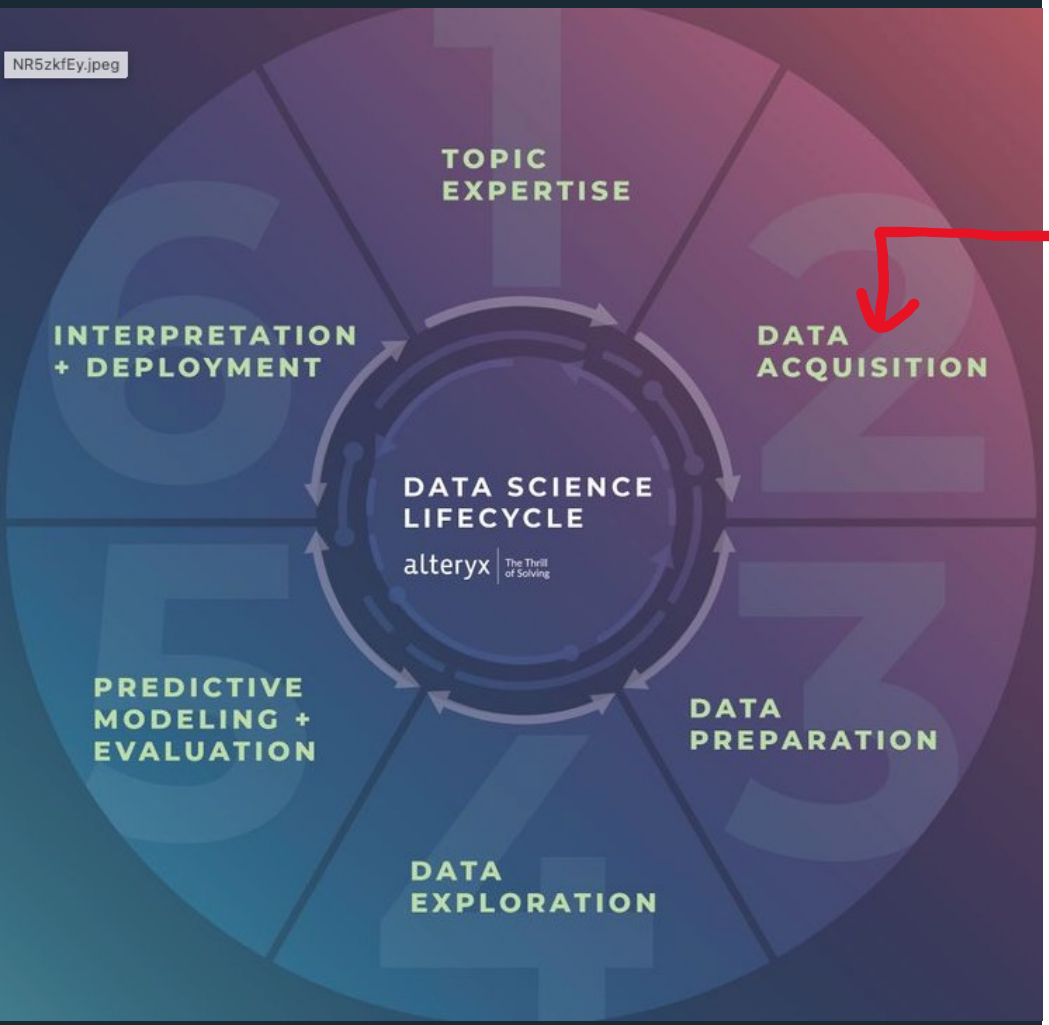
*"If we want it to play a positive role in tomorrow's world, it must be guided by human concerns"*

Feifei Li (Stanford's Human-Centered AI Institute)

# Human-AI teaming is important in healthcare domain, because

1. Complicated data and the requirement of extensive domain knowledge to process
2. Limited labeled datasets, leading to limited model performance
3. High-stakes decision-making
4. Complex contextual information and multiple factors
5. Ethical concerns

....




# Human-AI Teaming in Healthcare Data Analytics

- **Data collection**

  A case of virtual cognitive testing

- Model development

  AutoML + Visualization

- Decision-making

  A case of pediatric cardiac surgery

Data challenges:

- Lack of data
- Absence of labels
- Dirty data

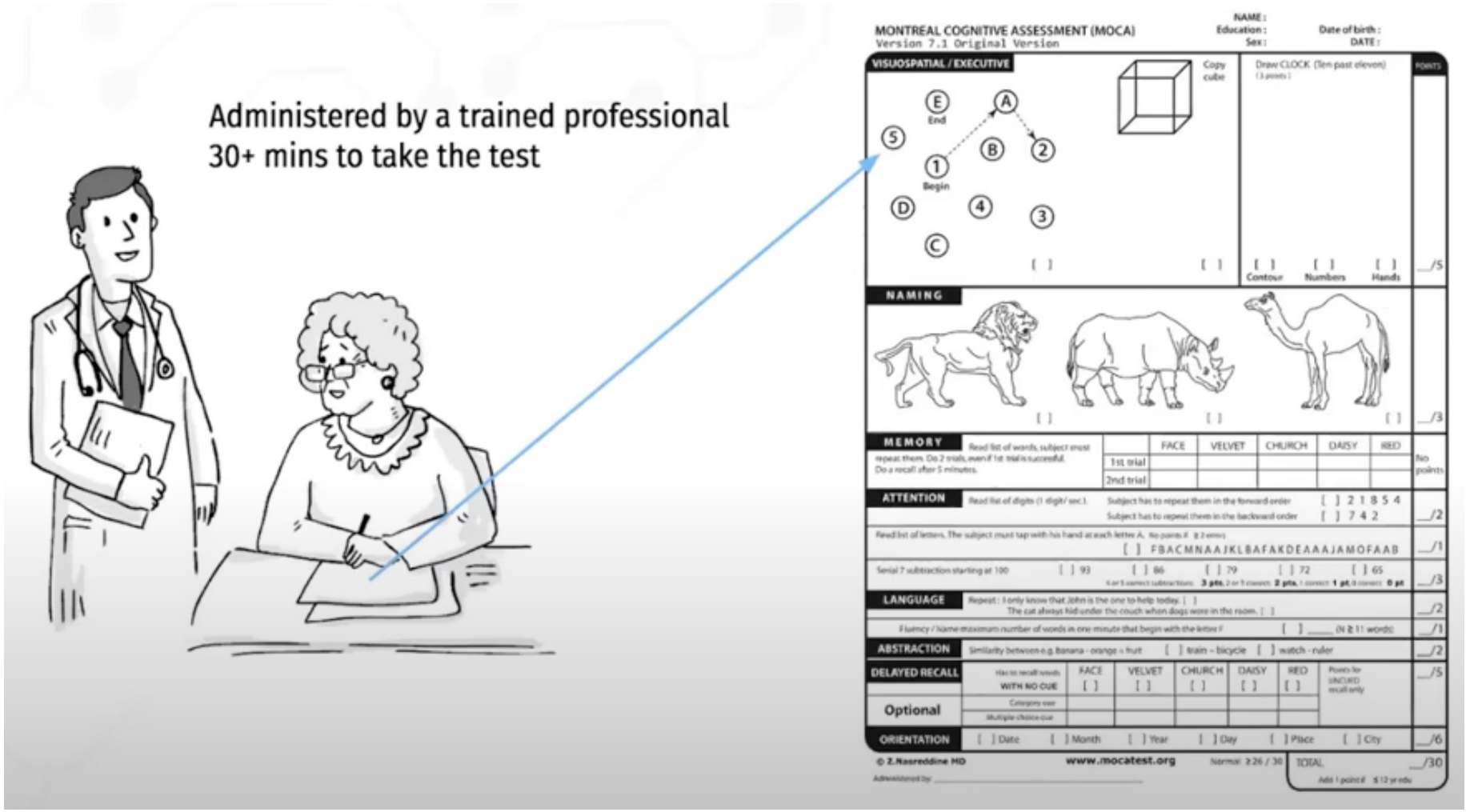
# The problem

The World Health Organization estimated that the total worldwide cost of dementia in 2018 was

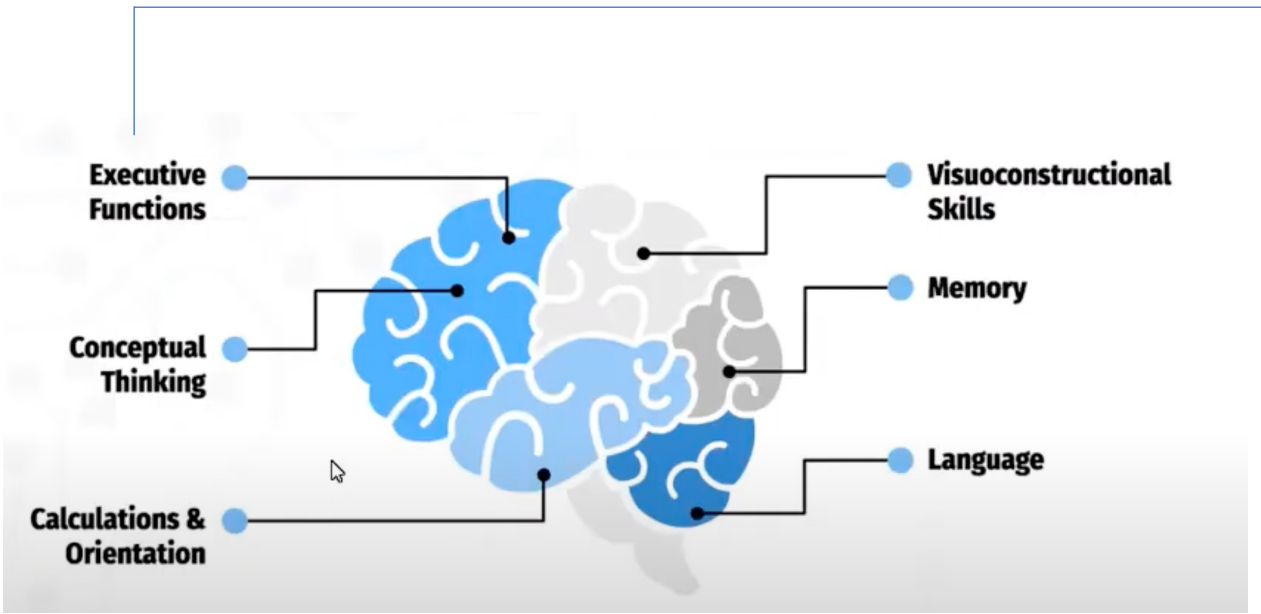
## $1 Trillion USD.

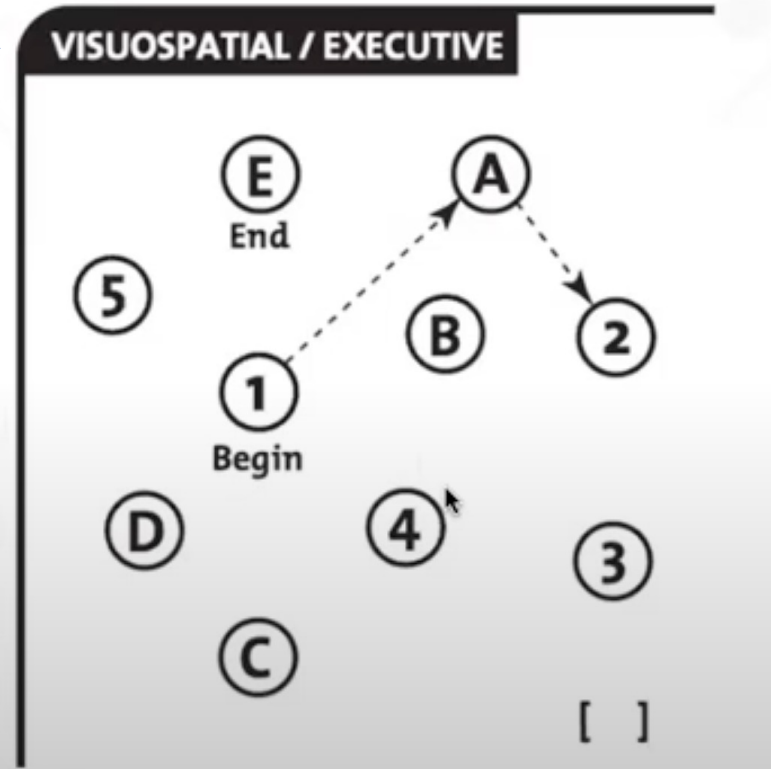
This figure is expected to rise to 2 trillion USD by 2030.

# How is the testing done?



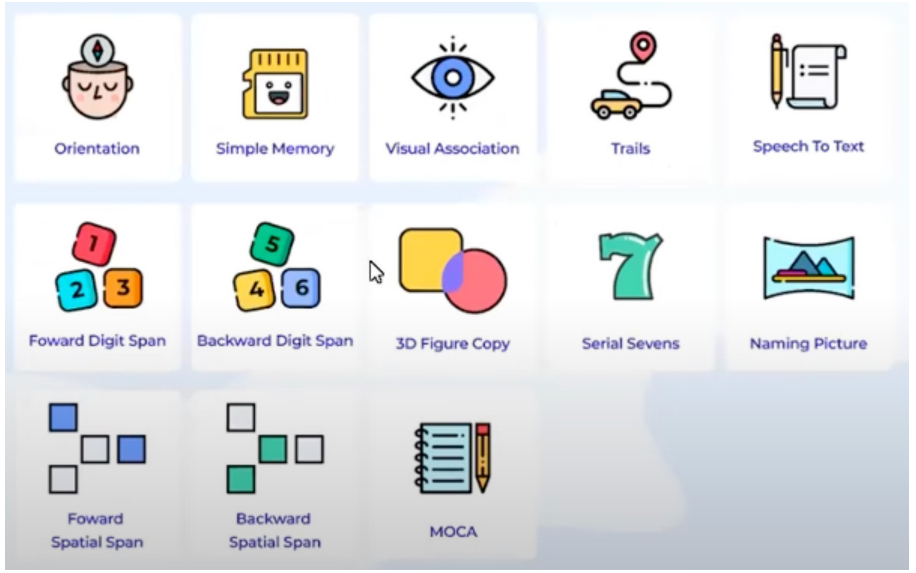
# Testing a variety of cognitive functions



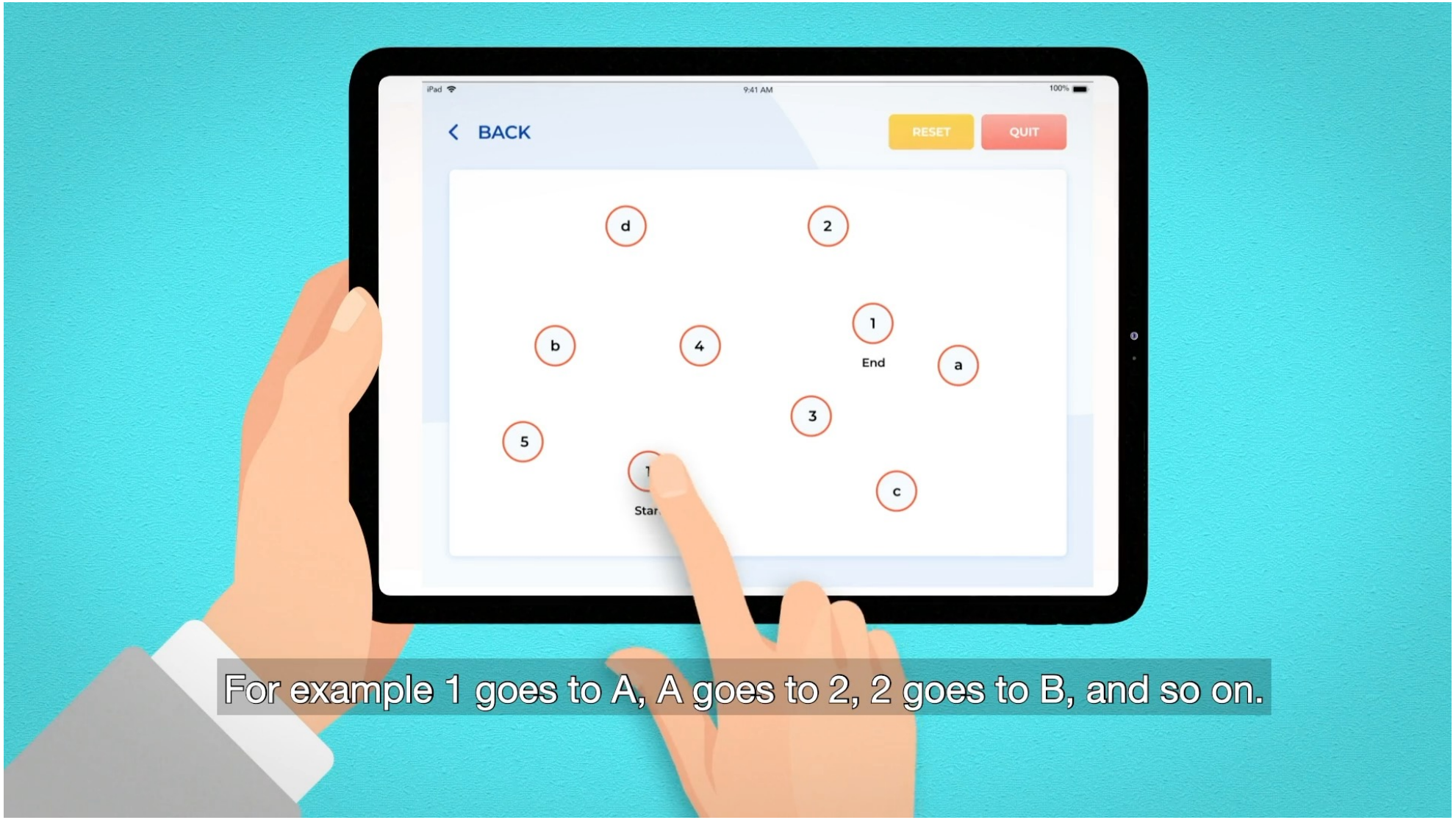
# IBOCA - A digital tool for the early detection of cognitive impairments

- Easily accessible, noninvasive, quick and comprehensive assessments.
- Accurate and fine-grained data collection.



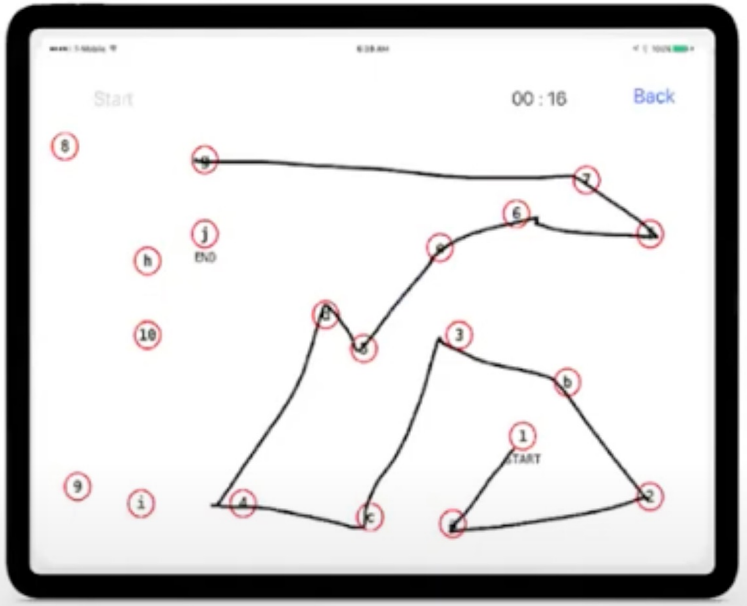

13 tests are currently included

https://dai.lids.mit.edu/iboca/



# The stats of data collected (very preliminary)


```
'Full Path': {'10.002012014389': {'status': 'Lineto',
  'x': '731.0',
  'y': '407.0'},
 '10.0178990364075': {'status': 'Lineto', 'x': '728.0', 'y': '406.0'},
 '10.034548997879': {'status': 'Lineto', 'x': '725.5', 'y': '405.0'},
 '10.0520520210266': {'status': 'Lineto', 'x': '723.0', 'y': '404.5'},
 '10.0680689811707': {'status': 'Lineto', 'x': '721.5', 'y': '403.5'},
 '10.084725022316': {'status': 'Lineto', 'x': '720.5', 'y': '402.5'},
 '10.1027799844742': {'status': 'Lineto', 'x': '719.5', 'y': '402.0'},
 '10.1192790269852': {'status': 'Lineto', 'x': '719.0', 'y': '401.5'},
 '10.1347559690475': {'status': 'Lineto', 'x': '718.5', 'y': '401.0'},
 '10.1523489952087': {'status': 'Lineto', 'x': '718.5', 'y': '400.5'},
 '11.2544059753418': {'status': 'Moveto', 'x': '710.5', 'y': '397.0'},
 '11.4330999851227': {'status': 'Lineto', 'x': '706.0', 'y': '396.5'},
 '11.4523929357529': {'status': 'Lineto', 'x': '705.0', 'y': '396.5'},
 '11.4692770242691': {'status': 'Lineto', 'x': '704.0', 'y': '396.5'},
 '11.4864829778671': {'status': 'Lineto', 'x': '702.5', 'y': '396.5'},
```



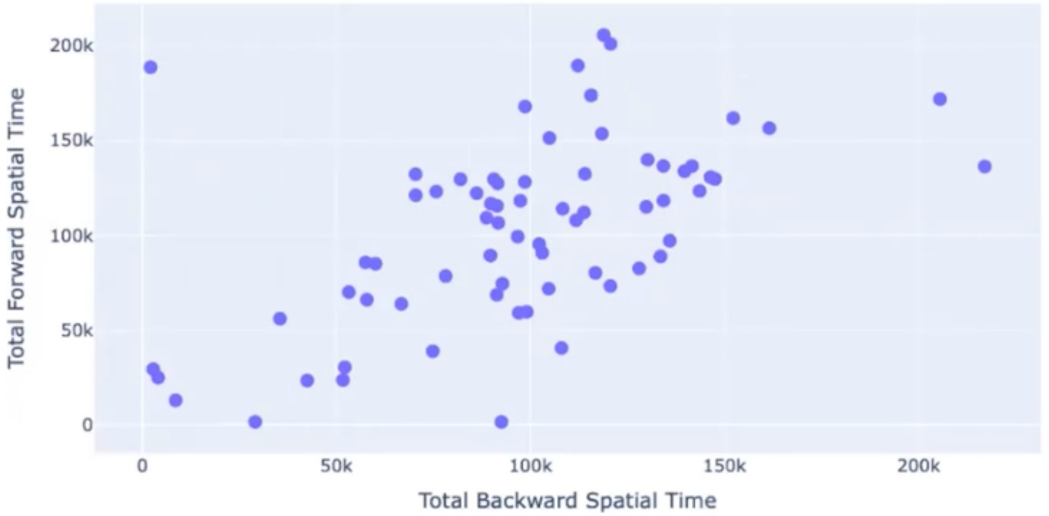
```
 '1': {'Correct': 'true', 'End': 1, 'Start': 0, 'Time (ms)': 9997},
 '2': {'Correct': 'true', 'End': 2, 'Start': 1, 'Time (ms)': 12414},
 '3': {'Correct': 'true', 'End': 3, 'Start': 2, 'Time (ms)': 13696}},
'Total Bubbles': 4
```

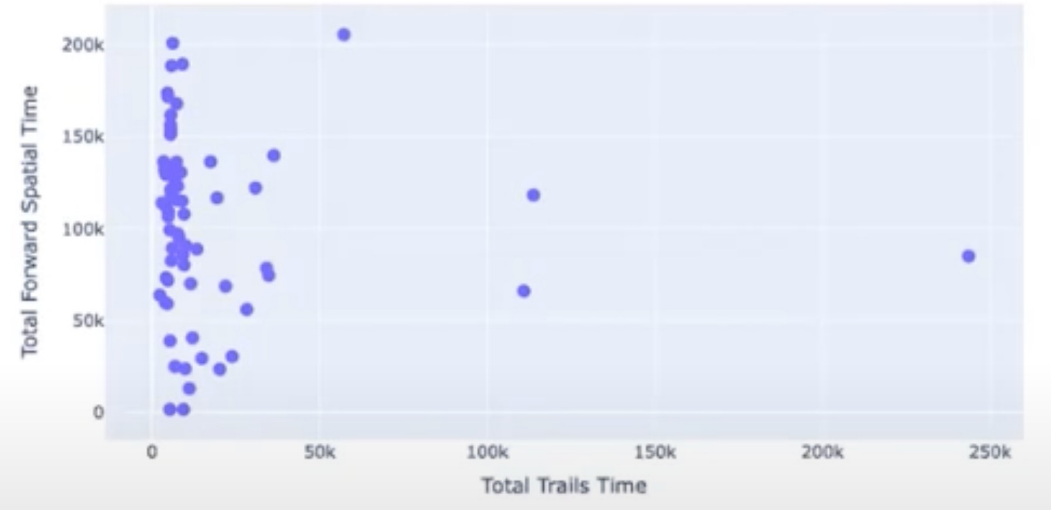

12 Patients tested remotely

# What has the data enabled?

**Conduct data analysis:**

- Is the time taken for these tests correlated?

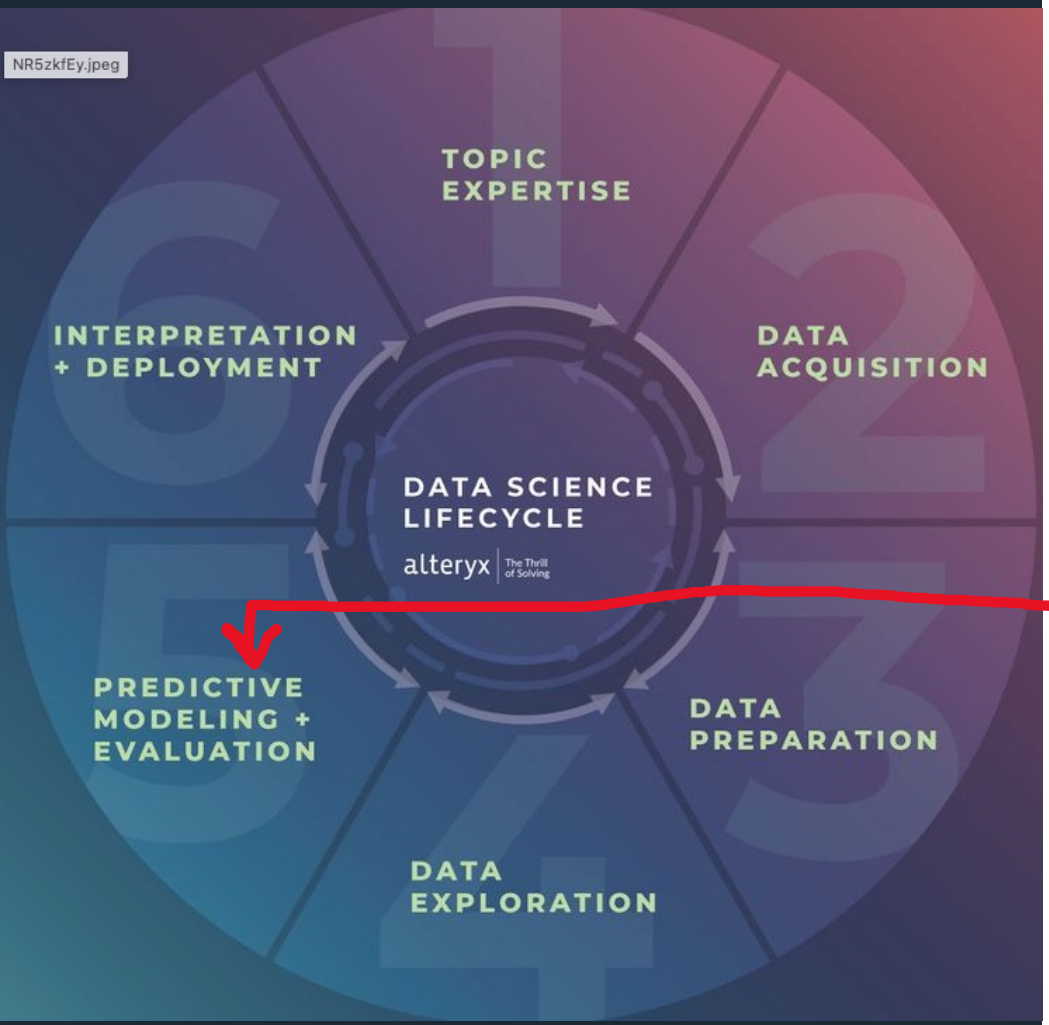


# What has the data enabled?

**Build AI/ML models to:**

- improve test accuracy
- create data-driven adaptive tests
- assist in clinical validation and decision-making
- and more…




# Human-AI Teaming in Healthcare Data Analytics

- Data collection

  A case of virtual cognitive testing

- **Model development**

  **AutoML + Visualization**

- Decision-making

  A case of pediatric cardiac surgery

# Electronic health record (EHR) dataset

scientific data

Explore content ⌄ About the journal ⌄ Publish with us ⌄

nature › scientific data › data descriptors › article

Open Access | Published: 24 May 2016

## MIMIC-III, a freely accessible critical care database

Alistair E.W. Johnson, Tom J. Pollard ✉, Lu Shen, Li-wei H. Lehman, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi & Roger G. Mark

*Scientific Data* **3**, Article number: 160035 (2016) | Cite this article

**205k** Accesses | **2834** Citations | **161** Altmetric | Metrics

# Electronic health record (EHR) dataset

scientific **data**

Explore content ⌄ About the journal ⌄ Publish with us ⌄

nature › scientific data › data descriptors › article

Open Access | Published: 24 May 2016

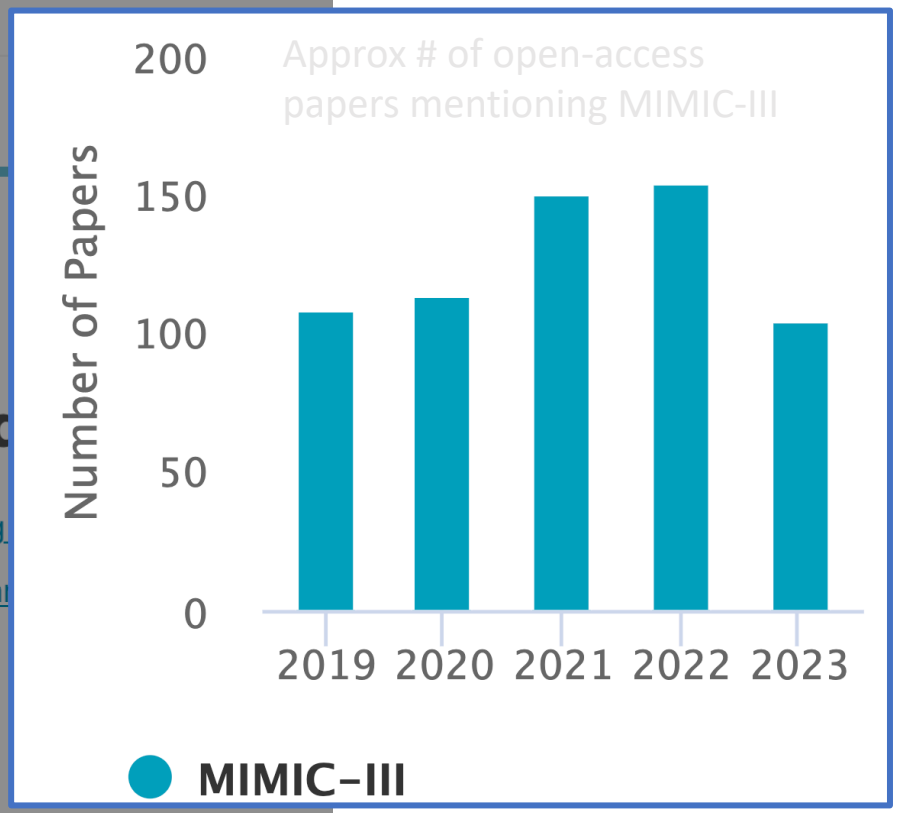
**MIMIC-III, a freely accessible critical care d**

Alistair E.W. Johnson, Tom J. Pollard ✉, Lu Shen, Li-wei H. Lehman, Mengling

Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi & Roger G. Mar

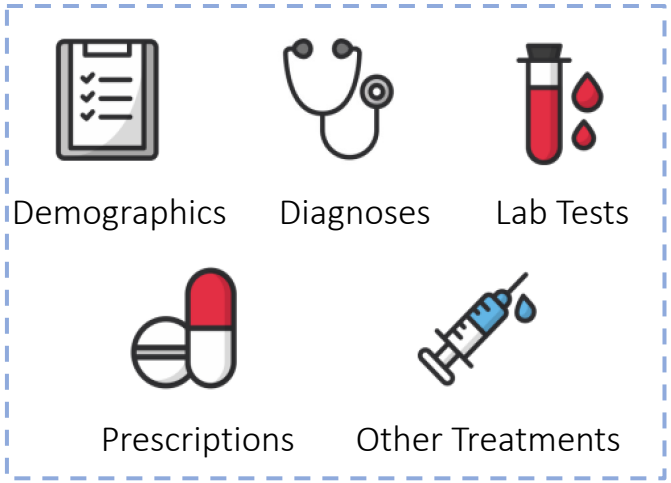
*Scientific Data* **3**, Article number: 160035 (2016) | Cite this article

**205k** Accesses | **2834** Citations | **161** Altmetric | Metrics



# How AI/ML models are being built?

EHR Database

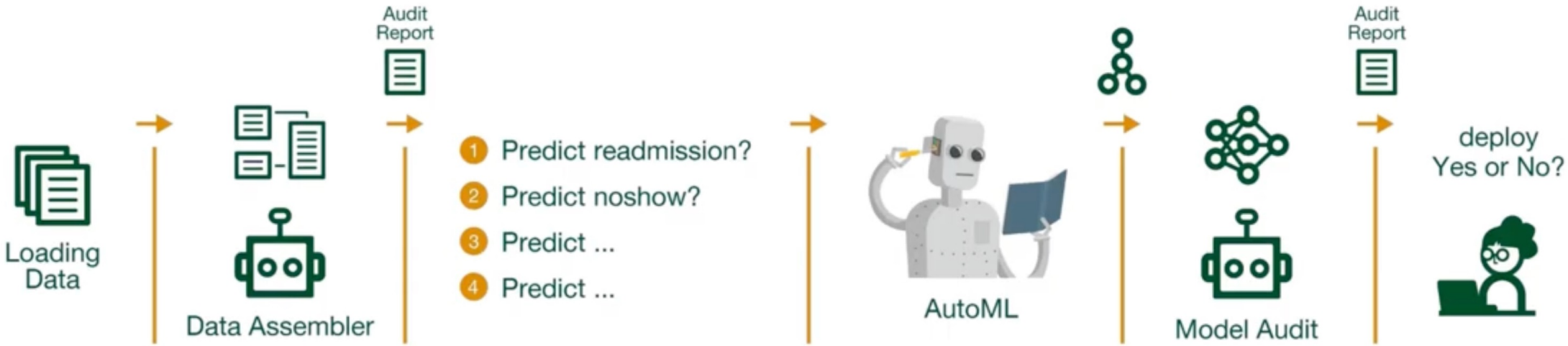



End User

ML Researcher/Engineer

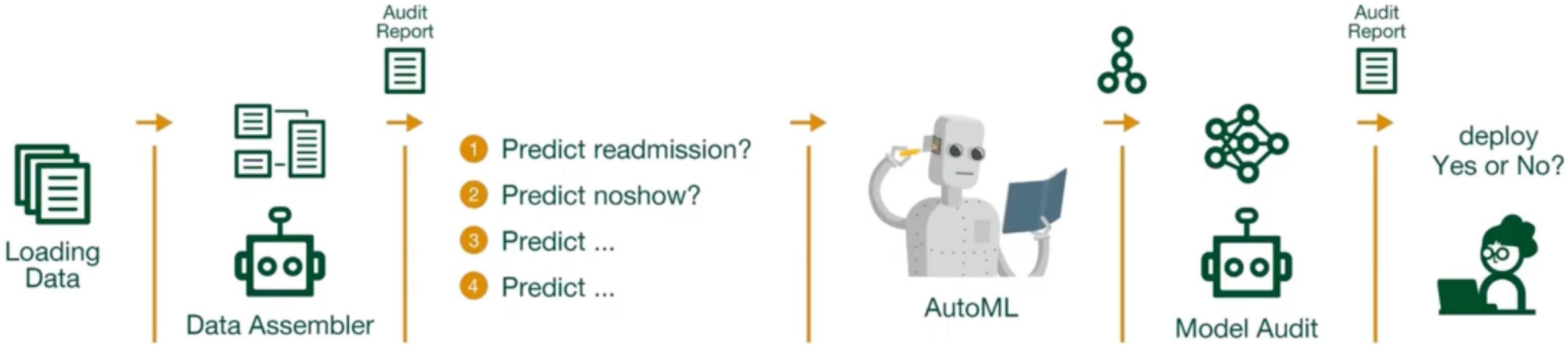
# Cardea Framework

Alnegheimish, Alrashed, Aleissa, Althobaiti, **Liu**, et al., Cardea: An Open Automated Machine Learning Framework for Electronic Health Records. DSAA 2020.



# Cardea Framework

Alnegheimish, Alrashed, Aleissa, Althobaiti, **Liu**, et al., Cardea: An Open Automated Machine Learning Framework for Electronic Health Records. DSAA 2020.




**Mortality**

Predict patient mortality from the point of admission.

**Readmission**

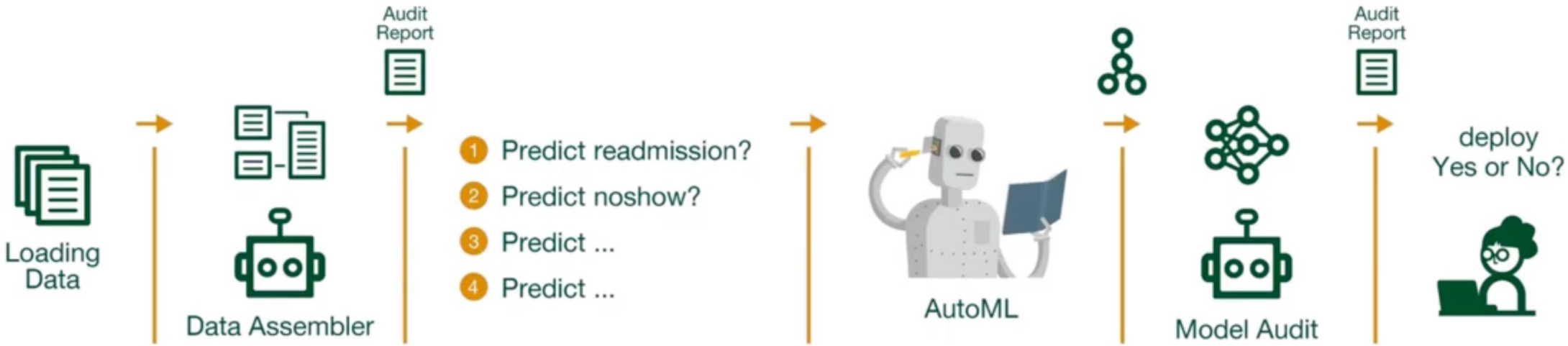
Predicts whether a patient will revisit the hospital after discharge.

**LOS ≥ 7**

Predict whether the patient will stay more than a week in the hospital.

# Cardea Framework

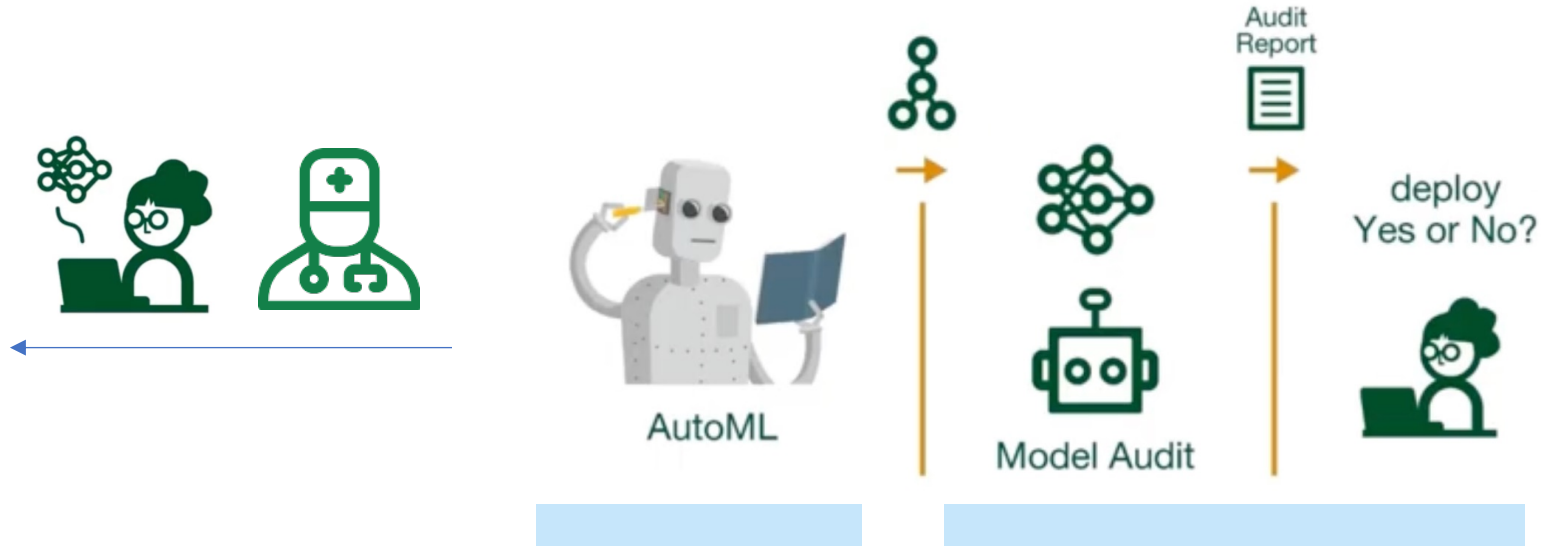
Alnegheimish, Alrashed, Aleissa, Althobaiti, **Liu**, et al., Cardea: An Open Automated Machine Learning Framework for Electronic Health Records. DSAA 2020.




| | F1 | | |
|---|---|---|---|
| | $\mu \pm \sigma$ | *best* | *best CL* |
| Mortality | $0.566 \pm 0.0529$ | 0.660 | XGB |
| Readmission | $0.540 \pm 0.0628$ | 0.635 | XGB |
| LOS $\geq 7$ | $0.519 \pm 0.1650$ | 0.789 | LR |

# The need to involve humans in the loop

Can the model be better?




| | *Cardea* | | |
|---|---|---|---|
| | $\mu \pm \sigma$ | *best* | *best CL* |
| Mortality | 0.566 ± 0.0529 | 0.660 | XGB |
| Readmission | 0.540 ± 0.0628 | 0.635 | XGB |
| LOS $\geq$ 7 | 0.519 ± 0.1650 | 0.789 | LR |

# Increase transparency and controllability in AutoML

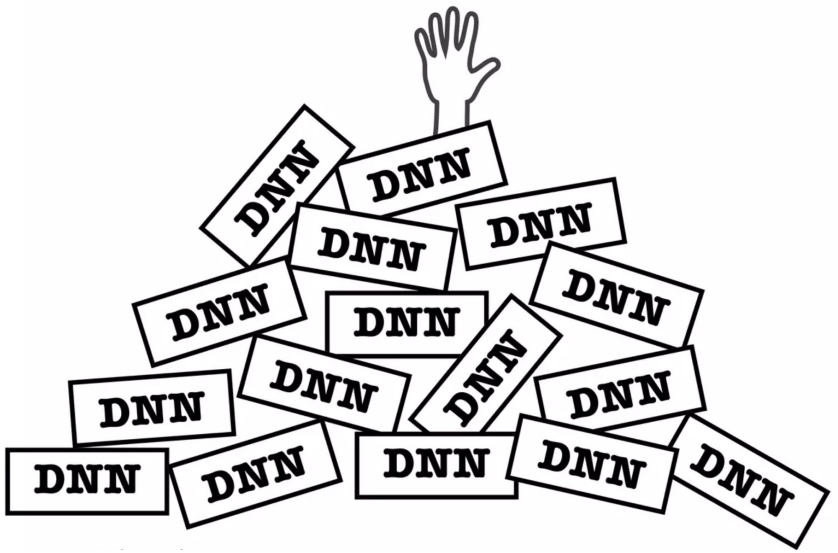
Random Forest

MLP

⋮

KNN

SVM



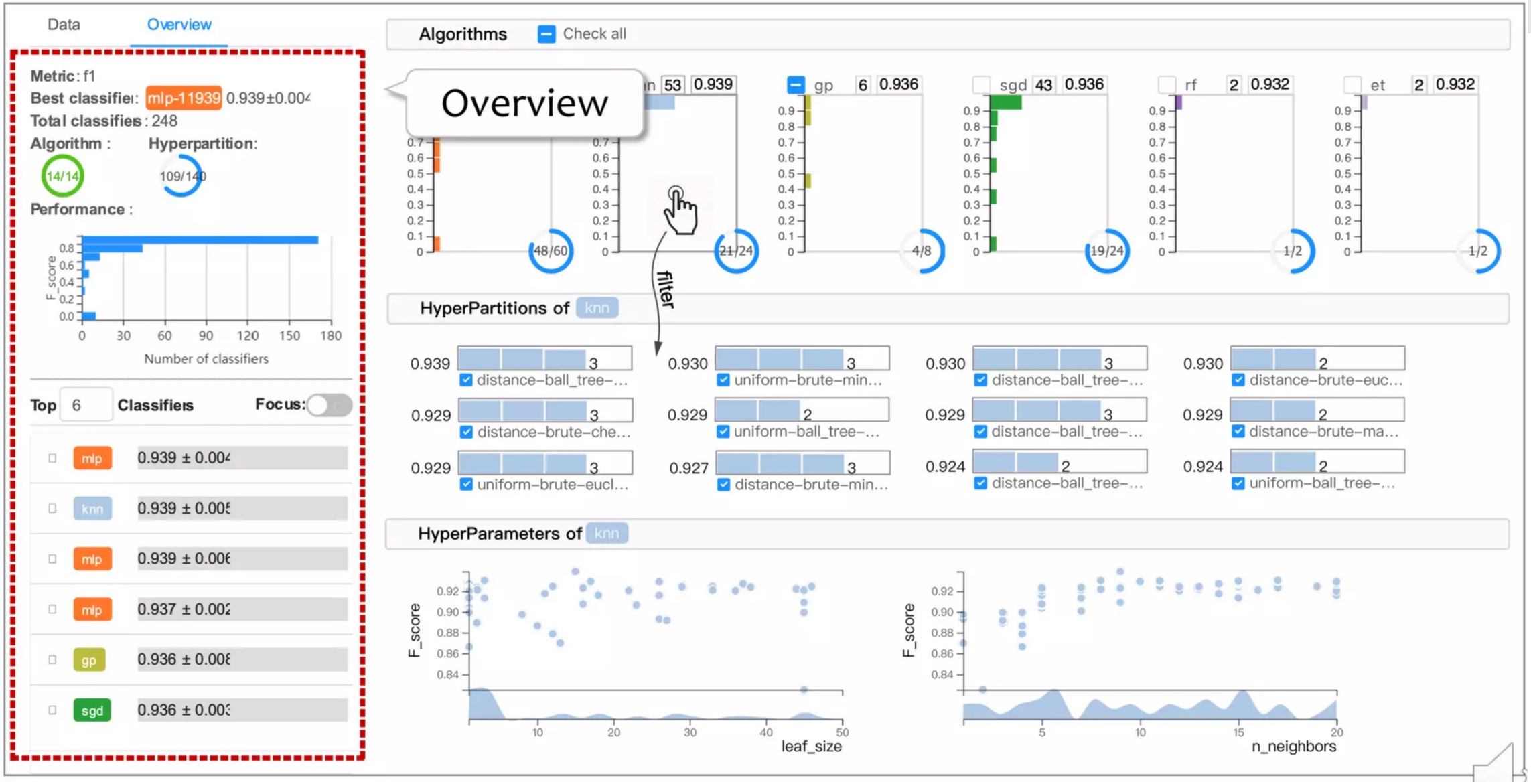

# layers = ?

# neurons = ?

⋮

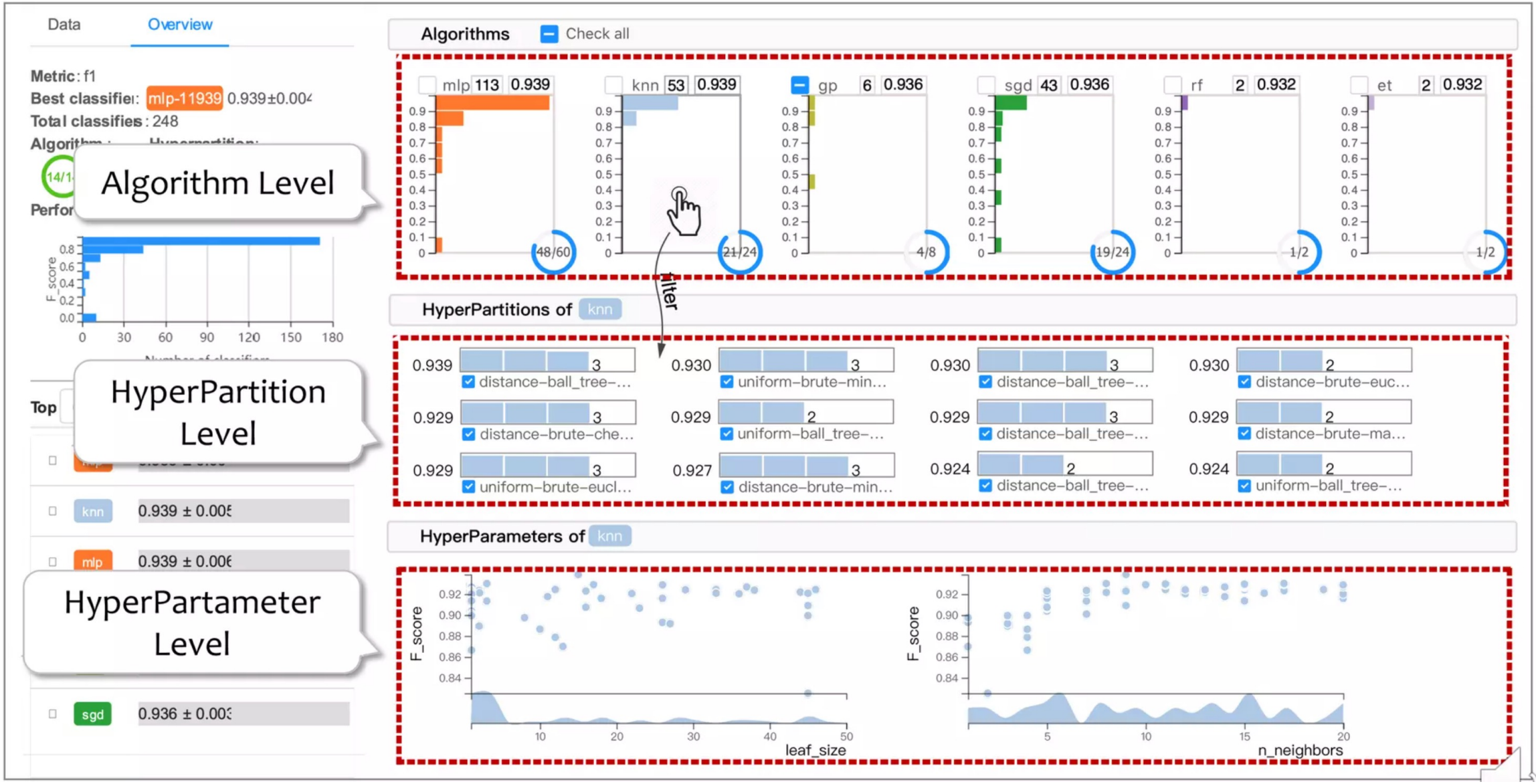
batch size =?

learning rate = ?



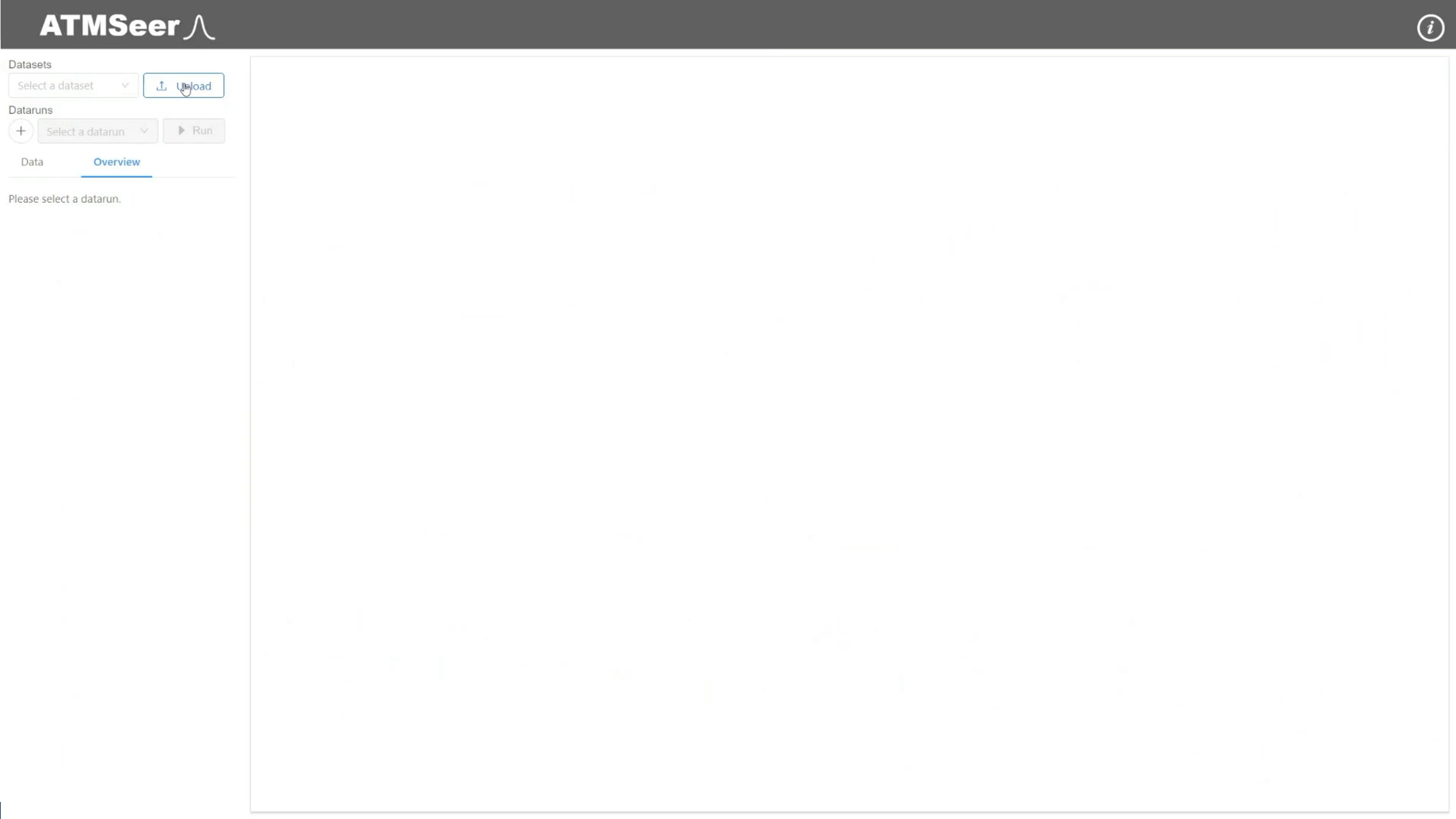

Wang, Ming, Jin, Shen, **Liu**, et al., ATMSeer: Increasing Transparency and Controllability in AutoMated Machine Learning. CHI 2019.

https://dai.lids.mit.edu/projects/atmseer/

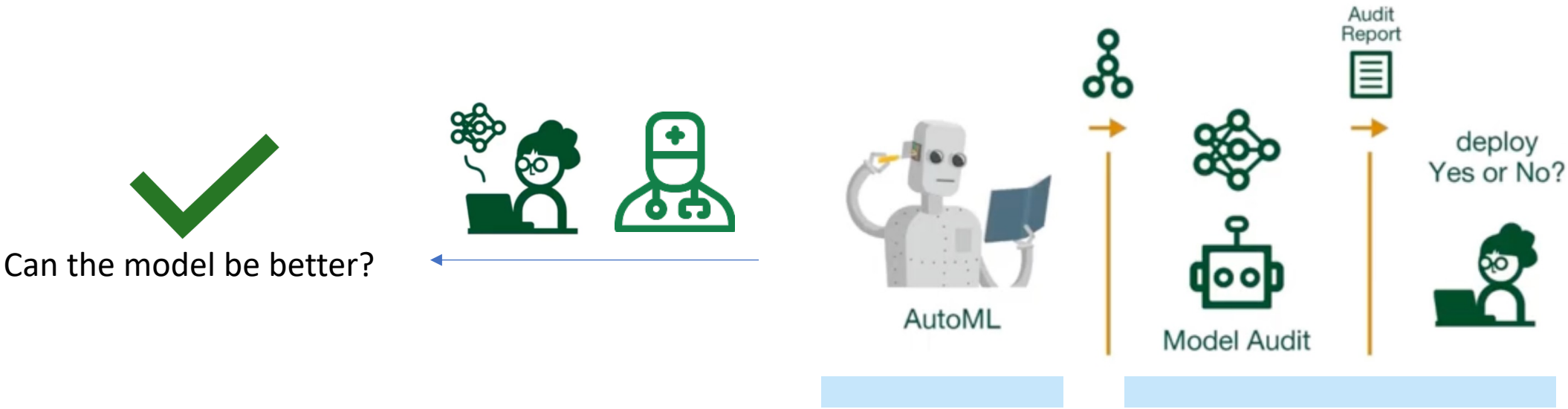



Wang, Ming, Jin, Shen, **Liu**, et al., ATMSeer: Increasing Transparency and Controllability in AutoMated Machine Learning. CHI 2019.

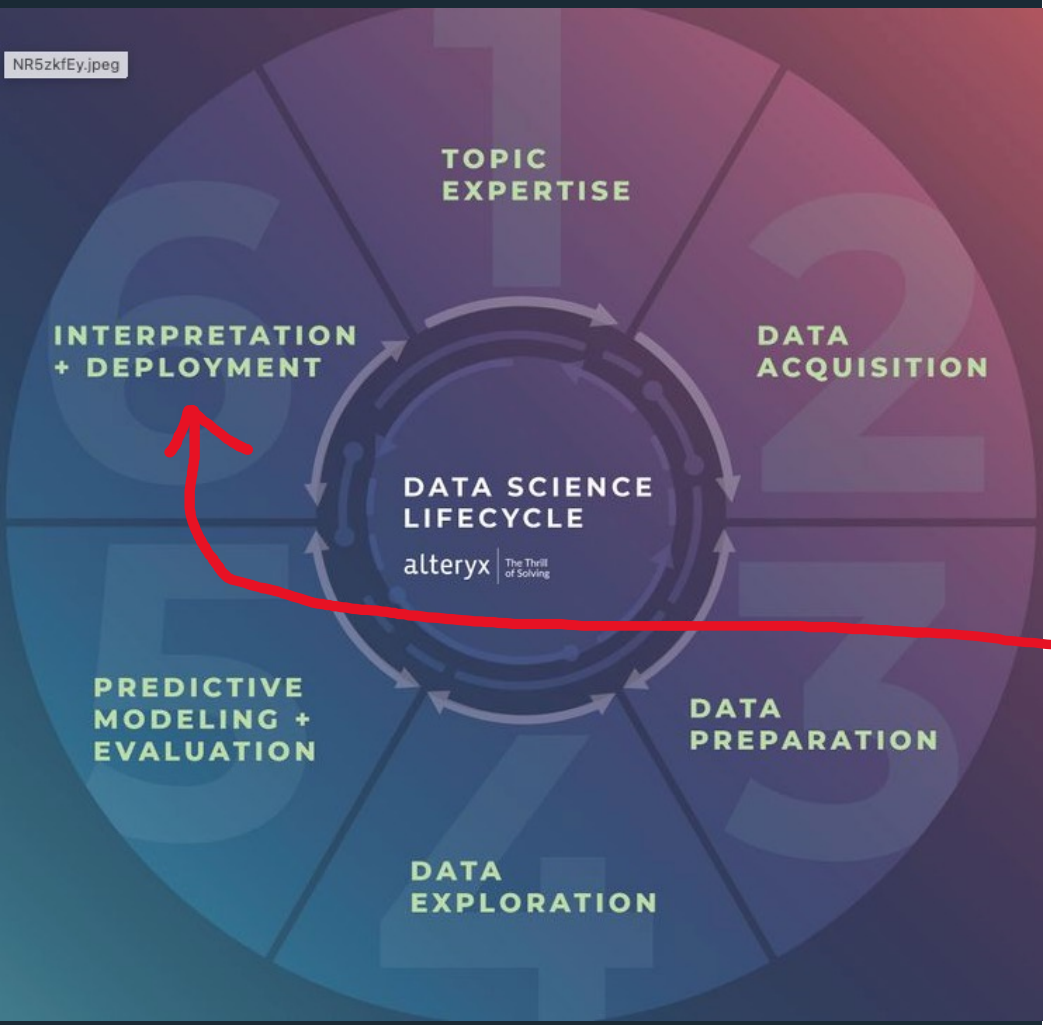
https://dai.lids.mit.edu/projects/atmseer/


ATMSeer
Datasets
Select a dataset
Upload
Dataruns
Select a datarun
Run
Data
Overview
Please select a datarun.

# The need to involve humans in the loop




| | *Cardea* | | |
|---|---|---|---|
| | $\mu \pm \sigma$ | *best* | *best CL* |
| Mortality | 0.566 ± 0.0529 | 0.660 | XGB |
| Readmission | 0.540 ± 0.0628 | 0.635 | XGB |
| LOS $\geq$ 7 | 0.519 ± 0.1650 | 0.789 | LR |




# Human-AI Teaming in Healthcare Data Analytics

- Data collection

  A case of virtual cognitive testing

- Model development

  AutoML + Visualization

- **Decision-making**

  **A case of pediatric cardiac surgery**

# Post-surgical complication prediction

Demographics

Diagnoses

Lab Tests

Prescriptions

Other Treatments

EHR Database

ML predicts five types of post-operative complications

Make treatment plan

# Research goal

- 6 experienced clinicians
- **Goal:** understand how clinicians expect to use ML models with feature explanations to support them make decisions.

# EHR Data – PIC
# A pediatric-specific intensive care database

scientific data

Explore content ⌄ About the journal ⌄ Publish with us ⌄

nature › scientific data › data descriptors › article

Data Descriptor | Open Access | Published: 13 January 2020

## PIC, a paediatric-specific intensive care database

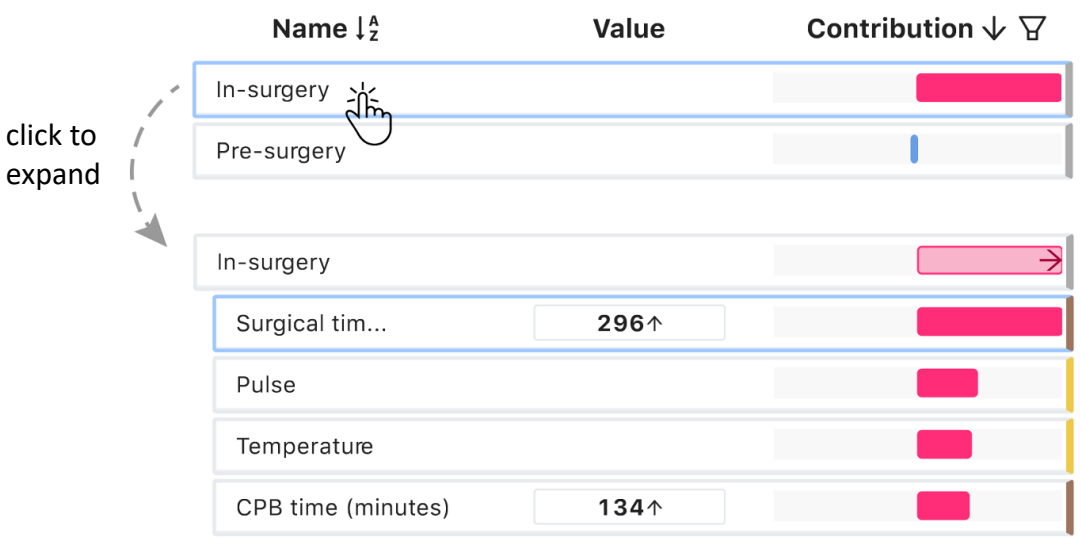
Xian Zeng, Gang Yu, Yang Lu, Linhua Tan, Xiujing Wu, Shanshan Shi, Huilong Duan, Qiang Shu ✉ & Haomin Li ✉

*Scientific Data* **7**, Article number: 14 (2020) | Cite this article

**8242** Accesses | **34** Citations | **47** Altmetric | Metrics

Follows the MIMIC-III format

# Inform the design #1




## Unfamiliarity with ML-engineered features

*"We don't often use statistical values like Trend or Standard Deviation (SD) in our work."*
(Clinician-P5)

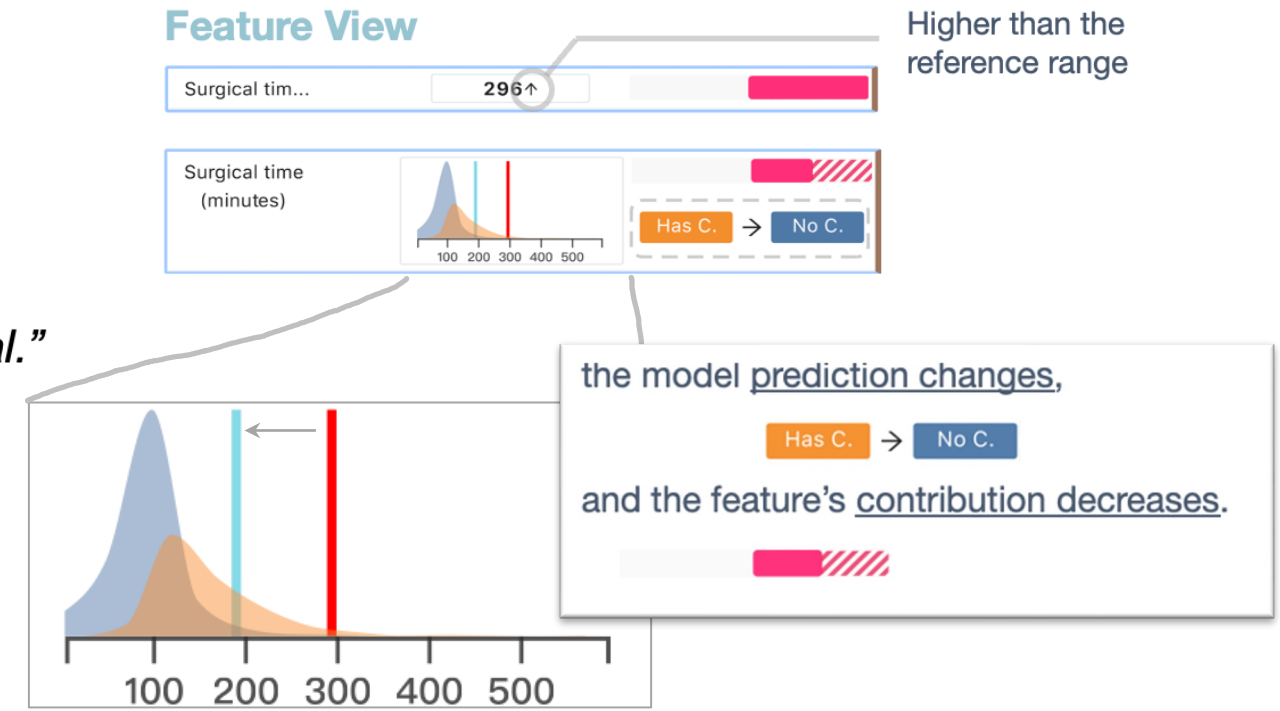
*"Users preferred to see related features together or how features may interact."*
(Wang et al., 2019)

Cheng, **Liu**, et al., VBridge: Connecting the Dots Between Features and Data to Explain Healthcare Models, TVCG (VIS'21). Best Paper Honorable Mention.

# Inform the design #2

## Aligning with evidence

*"I want to see the records from other patients and whether the target patient's record values are within the 95% Confidence Interval."*
(Clinician-P1, Q7)



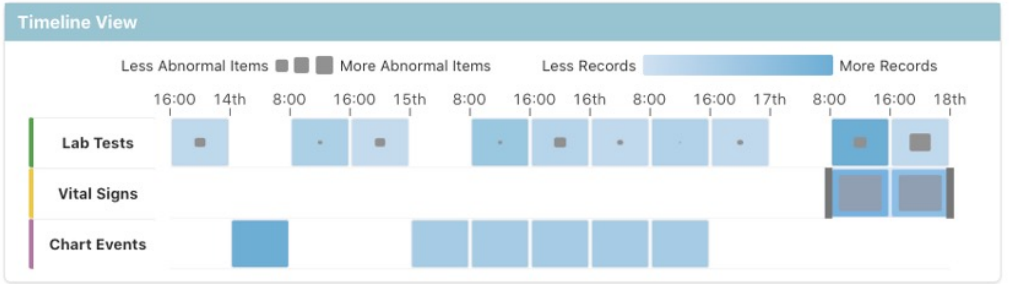

Cheng, **Liu**, et al., VBridge: Connecting the Dots Between Features and Data to Explain Healthcare Models, TVCG (VIS'21). Best Paper Honorable Mention.

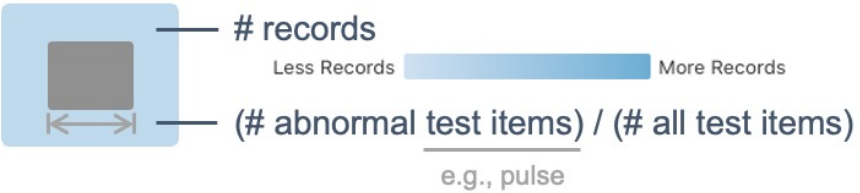
# Inform the design #3

## Connecting to original records

*"I would also like to see other types of records from the patients to gain a comprehensive understanding of the patients' situation."* (Clinician-P2, Q7)

**Timeline View: Overview of the longitude health records**



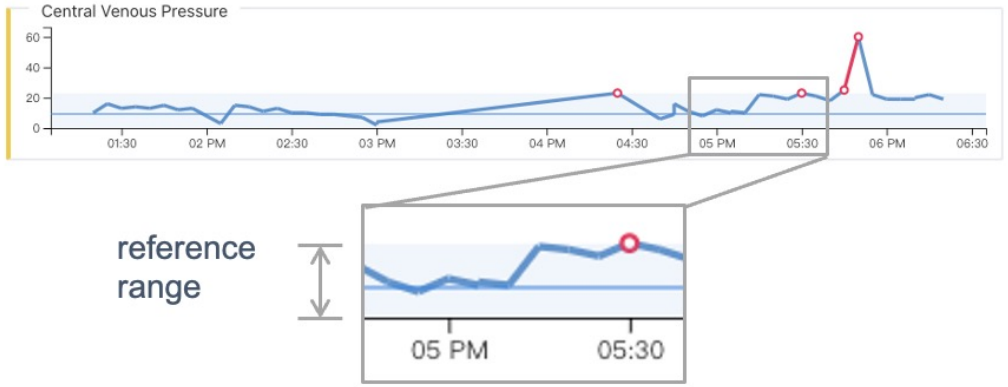




Cheng, **Liu**, et al., VBridge: Connecting the Dots Between Features and Data to Explain Healthcare Models, TVCG (VIS'21). Best Paper Honorable Mention.

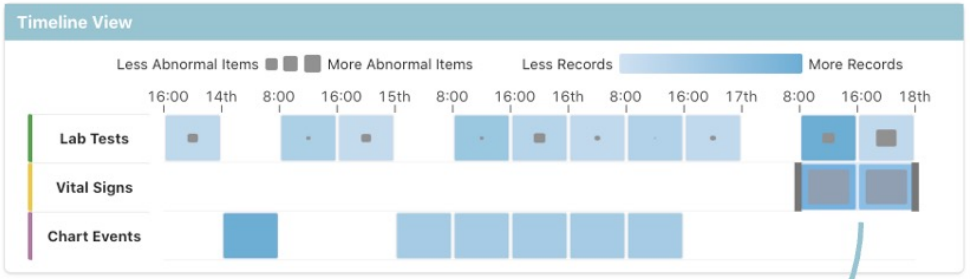
# Inform the design #3

## Connecting to original records

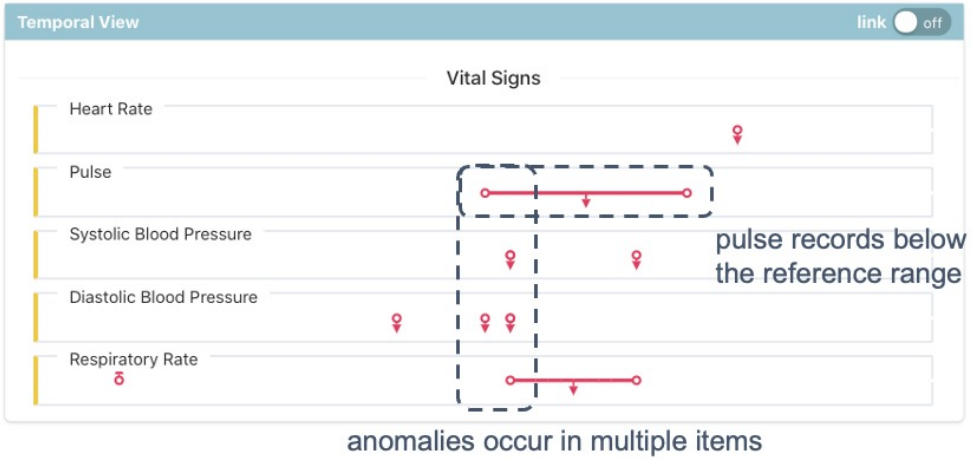
*"I would also like to see other types of records from the patients to gain a comprehensive understanding of the patients' situation."* (Clinician-P2, Q7)




### Timeline View: Overview of the longitude health records



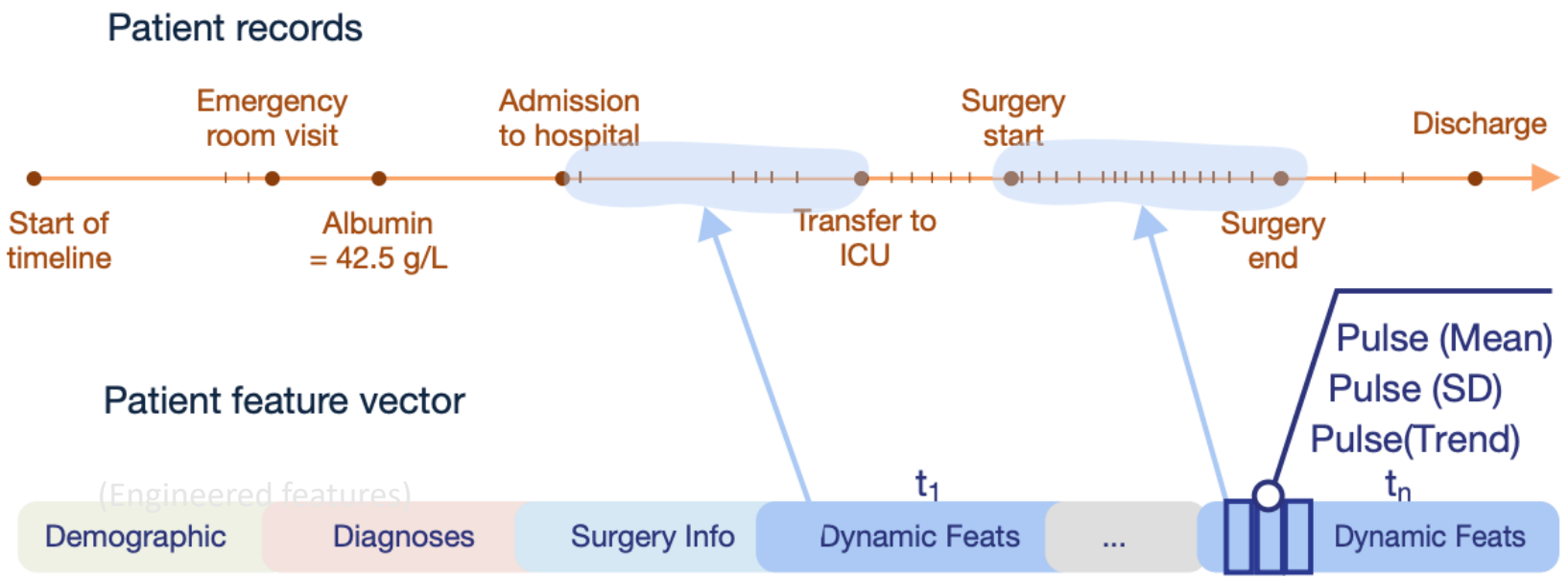

### Temporal View: Details of health records time series




Cheng, **Liu**, et al., VBridge: Connecting the Dots Between Features and Data to Explain Healthcare Models, TVCG (VIS'21). Best Paper Honorable Mention.

# Inform the design #3

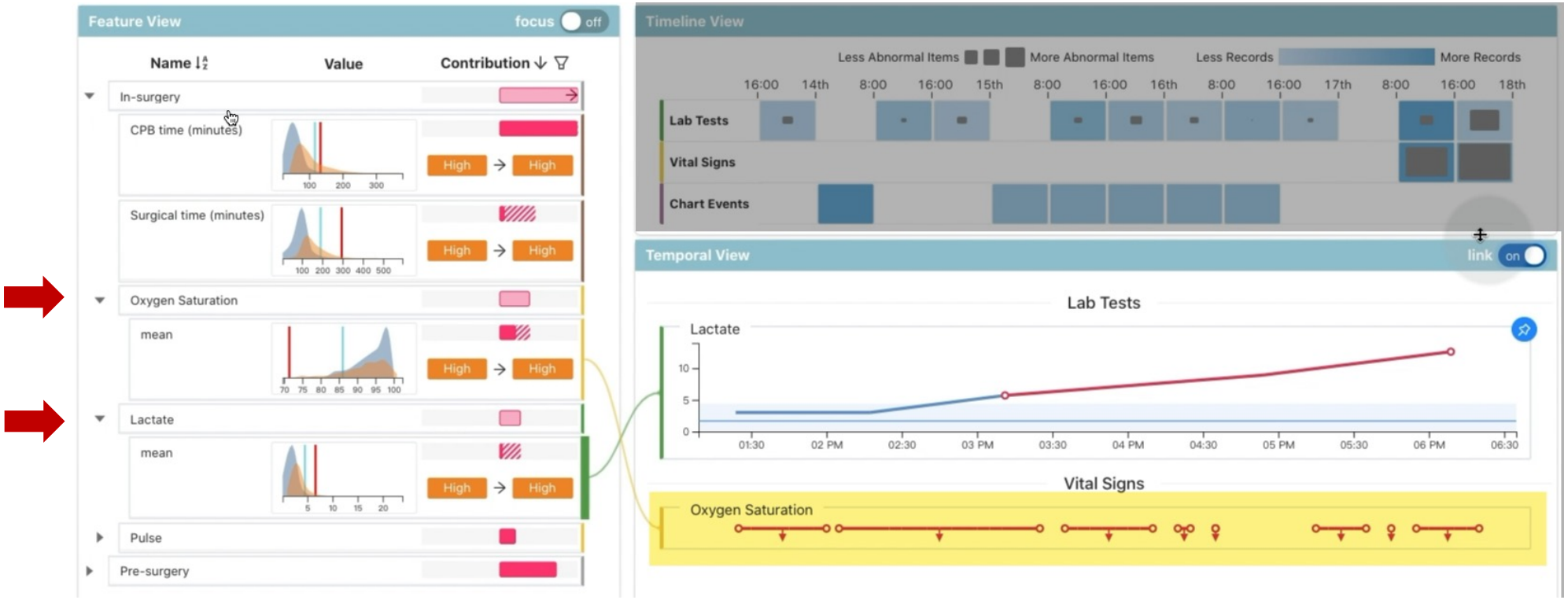
- Connect the dots between features and data




Cheng, **Liu**, et al., VBridge: Connecting the Dots Between Features and Data to Explain Healthcare Models, TVCG (VIS'21). Best Paper Honorable Mention.

# Inform the design #3

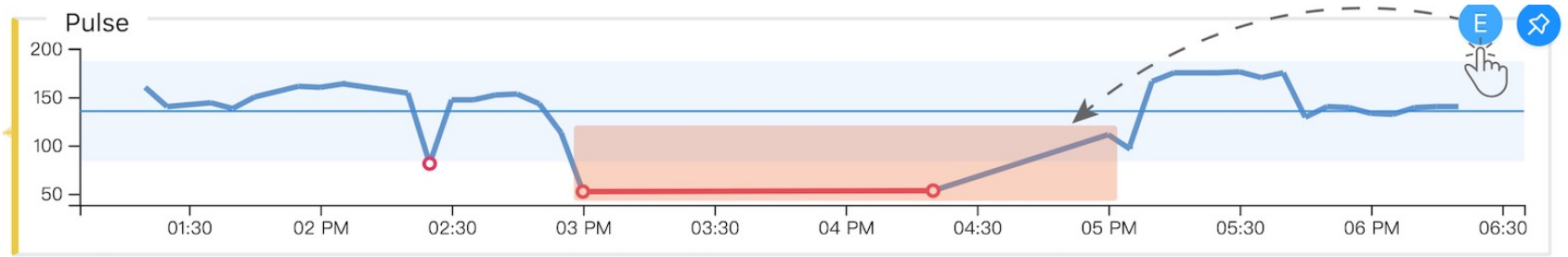
- Connect the dots between features and data




Cheng, **Liu**, et al., VBridge: Connecting the Dots Between Features and Data to Explain Healthcare Models, TVCG (VIS'21). Best Paper Honorable Mention.

# Inform the design #3

- Connect the dots between features and data




Influential time periods to feature values

Cheng, **Liu**, et al., VBridge: Connecting the Dots Between Features and Data to Explain Healthcare Models, TVCG (VIS'21). Best Paper Honorable Mention.

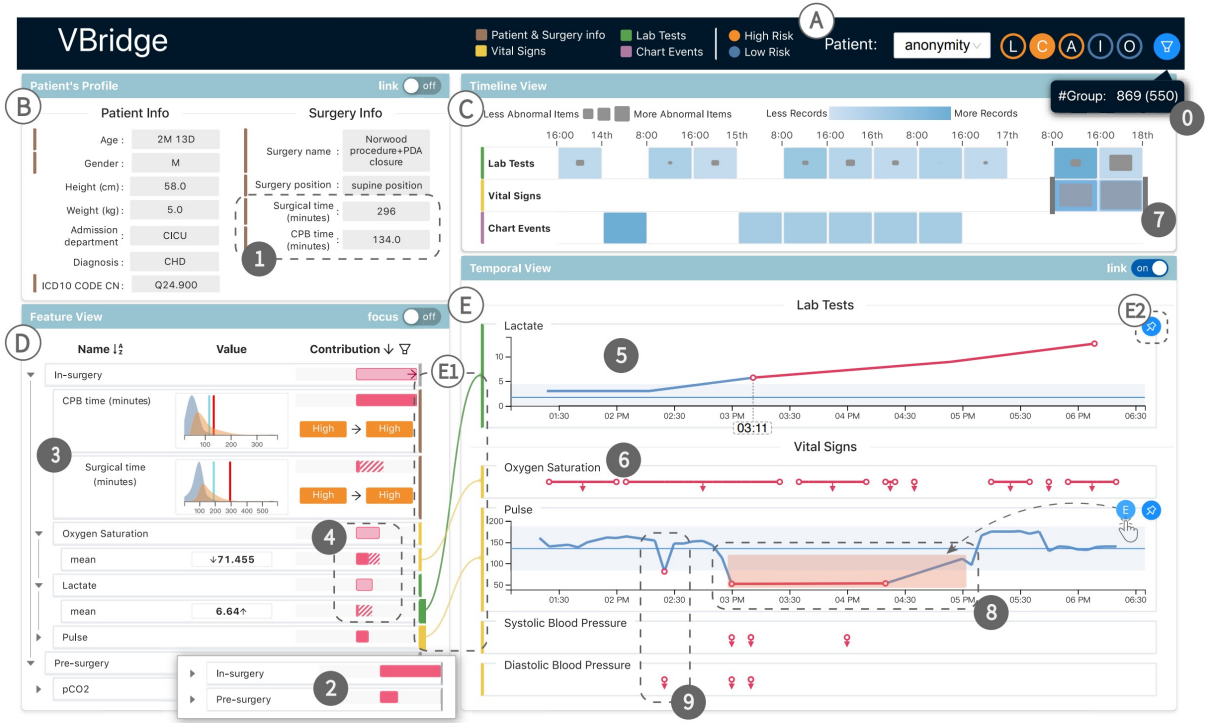
# Two analysis workflows

**Forward**

Data → Prediction → Explanations (Features)

**Backward**

Prediction → Explanations (Features) → Data




Cheng, **Liu**, et al., VBridge: Connecting the Dots Between Features and Data to Explain Healthcare Models, TVCG (VIS'21). Best Paper Honorable Mention.

# Recap ILOs

By the end of this class, you should be able to

1. **Describe** what is the data science life cycle
2. **Recognize** the importance of visualization-empowered human AI teaming in such a life cycle
3. **Explain** three key scenarios how human AI teaming can help with healthcare data analytics

# Questions and Feedback

Dongyu Liu

dyuliu@ucdavis.edu

dongyu.tech

UCDAVIS COMPUTER SCIENCE