Machine Learning and Rare Disease Research:

a boost for innovation

Jineta Banerjee

Principal Scientist, Sage Bionetworks

SageBionetworks

In the talk today:

- Where we are today and where we would like to go

- Machine learning and AI in rare disease today
  - Definitions
  - Challenges and Promises
    - Patients and therapies
    - Caregivers

- Looking to the future
  - Challenges and Promises:
    - Multi-modal data from patients
    - Digital twin technology
    - Foundation models in medicine

Present day’s headlines for rare diseases

Government of Canada | Gouvernement du Canada

Search Canada.ca

MENU

Canada.ca › Departments and agencies › Health Canada

Government of Canada improves access to affordable and effective drugs for rare diseases

From: Health Canada

News release

March 22, 2023 | Ottawa, Ontario | Health Canada

EDITION IN | SEATTLE 45°F

THE TIMES OF INDIA SUBSCRIBE TO TOI+

India | Saving Our Stripes | Times Evoke | Maharashtra | Delhi | Karnataka | Tamil Nadu | Telangana | Uttar Pradesh

NEWS / INDIA NEWS / Why Parents Of Children With Rare Genetic D...

TRENDING | Vizianagaram Train Accident | Kerala Blast News | Manish Sisodia | LRSAM

Why parents of children with rare genetic diseases have turned activists

Himanshi Dhawan / TNN / Updated: Oct 29, 2023, 01:15 IST

SHARE | AA | FOLLOW US

nature

Explore content | About the journal | Publish with us | Subscribe

nature › news › article

NEWS | 06 October 2023 | Correction 06 October 2023 | Correction 09 October 2023

Gene therapies for rare diseases are under threat. Scientists hope to save them

As industry steps aside, scientists seek innovative ways to make sure expensive treatments can reach people who need them.

What do we want the headlines to be?

Early diagnosis of rare disease X now possible

New and improved treatment now available for rare disease Y

FDA approves the first drug that targets rare disease Z

But we are not there yet!

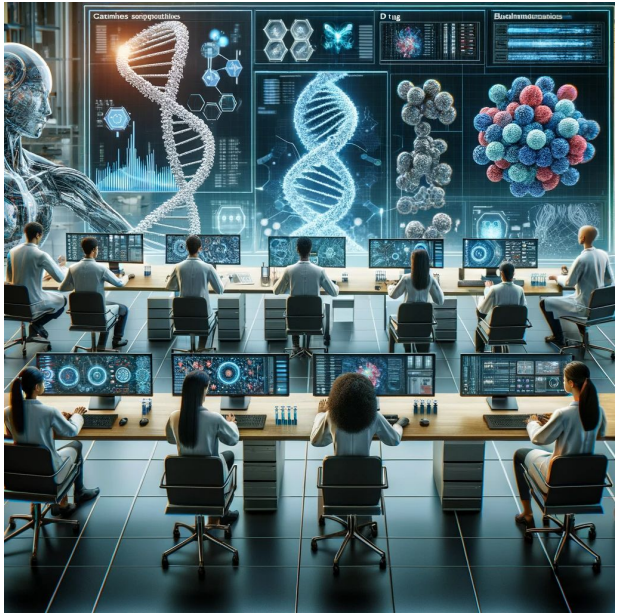
What do we need to get there?

Molecular, Pharmaceutical, and Clinical Research

Images courtesy: DALL-E3

What do we need to get there?

Molecular, Pharmaceutical, and Clinical Research



Accelerate biomedical discovery using computational methods

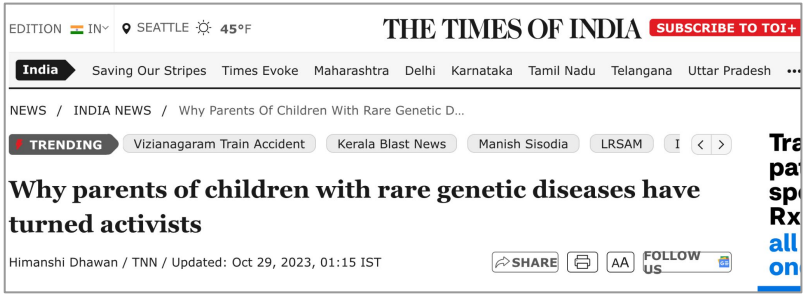
Images courtesy: DALL-E3

What do we need to get there?

Traditional Machine Learning (ML)
and
Artificial Intelligence (AI)

Images courtesy: DALL-E3

How can AI or ML help people living with rare diseases?



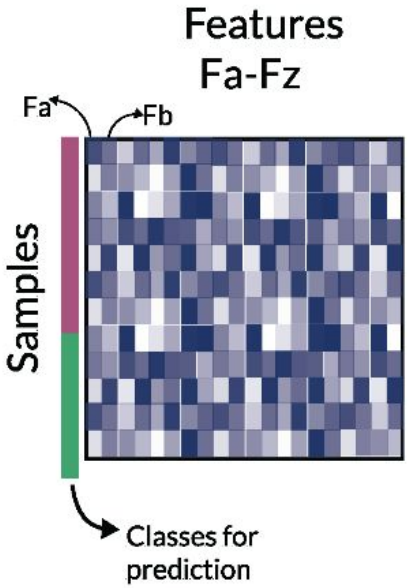
Images courtesy: DALL-E3

Traditional machine learning (ML) in rare disease

- Biomarkers
- Improved therapeutic targets

Images courtesy: DALL-E3

Definitions



TRADITIONAL ML TASK:

- Identifying a disease sample based on patterns in given data

FEATURES:

- individual measurable property or characteristic of a phenomenon

SAMPLES:

- Individual data points

CLASS / LABEL:

- Disease type

TRAINING DATA:

- Data to train the model to identify disease sample

TESTING DATA:

- Data previously unseen by the model

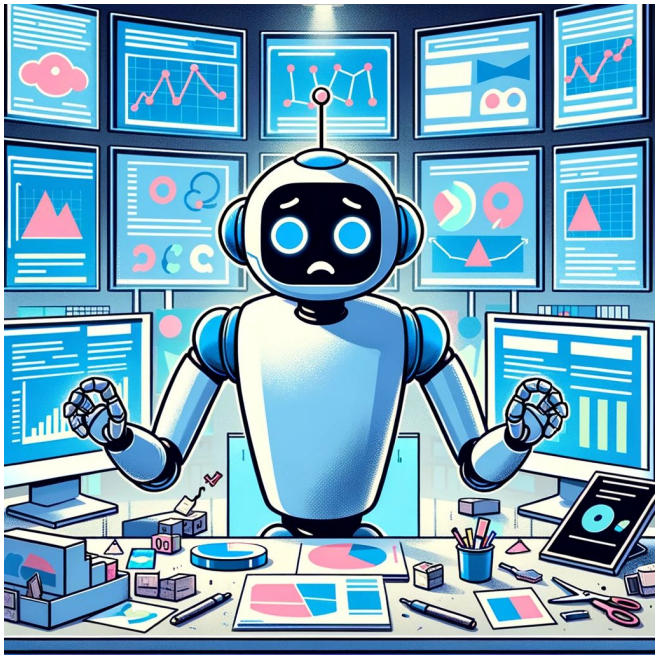
Bishop, 2006, *Pattern recognition and machine learning*.

ML in rare disease faces many challenges

- Data
- Method

Images courtesy: DALL-E3

Data challenges for ML in rare disease

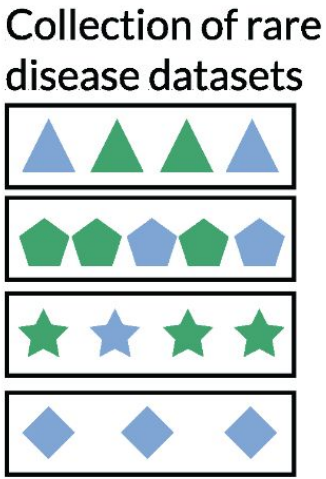


Images courtesy: DALL-E3

Machine learning needs large datasets

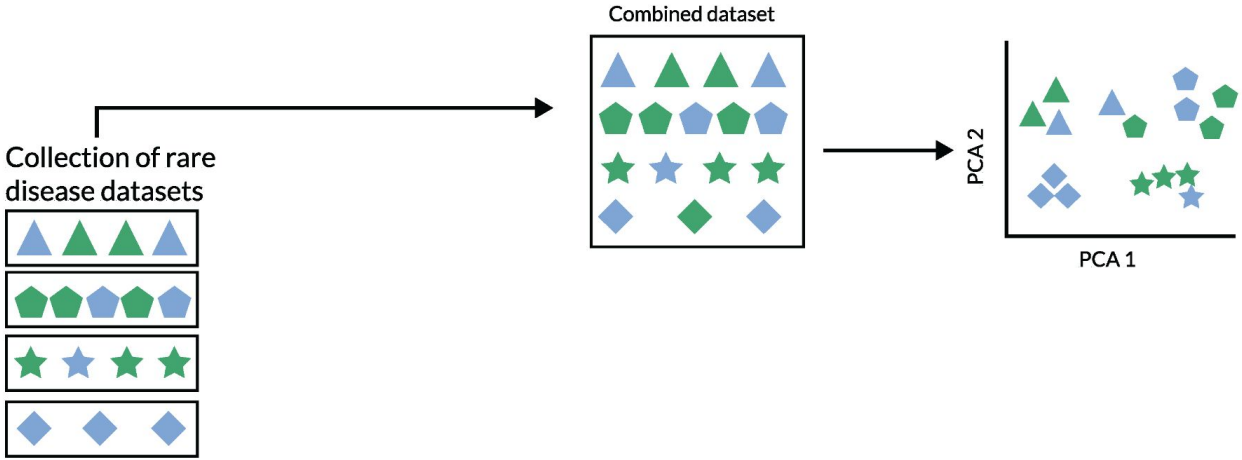
- Biomedical data contain many features
- To detect patterns effectively, algorithms need access to many samples
- “Curse of dimensionality”

Rare disease, by definition, has few patients or samples



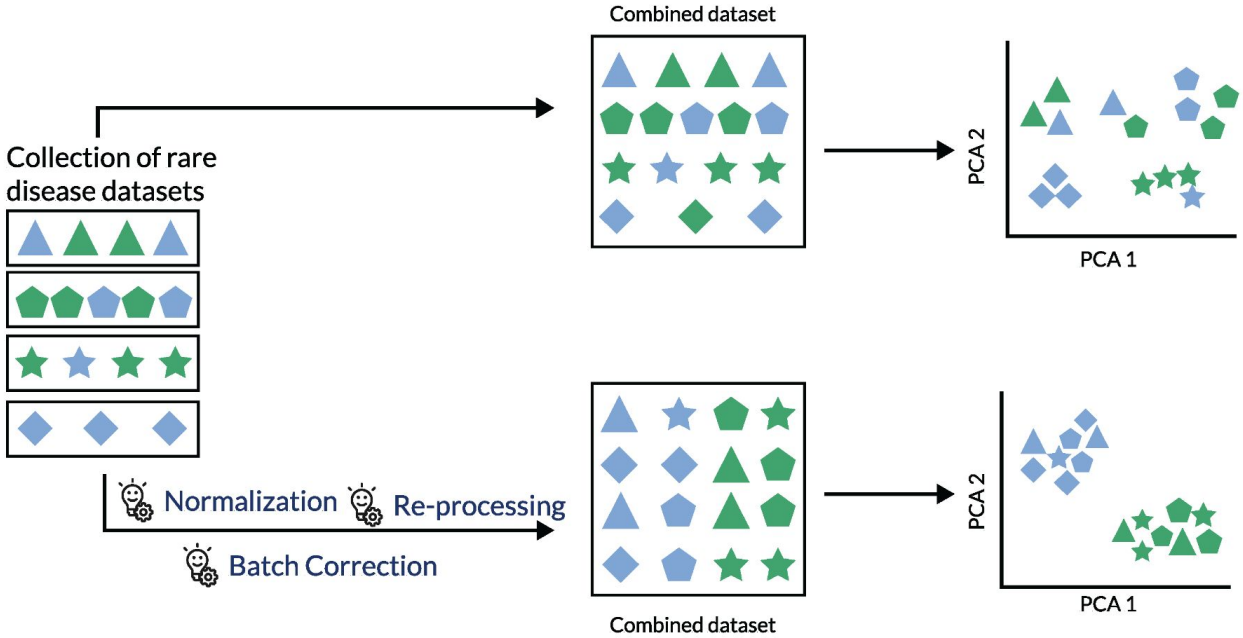
Legend: Color = disease, Shape = dataset origin

Combine multiple datasets to make one big dataset



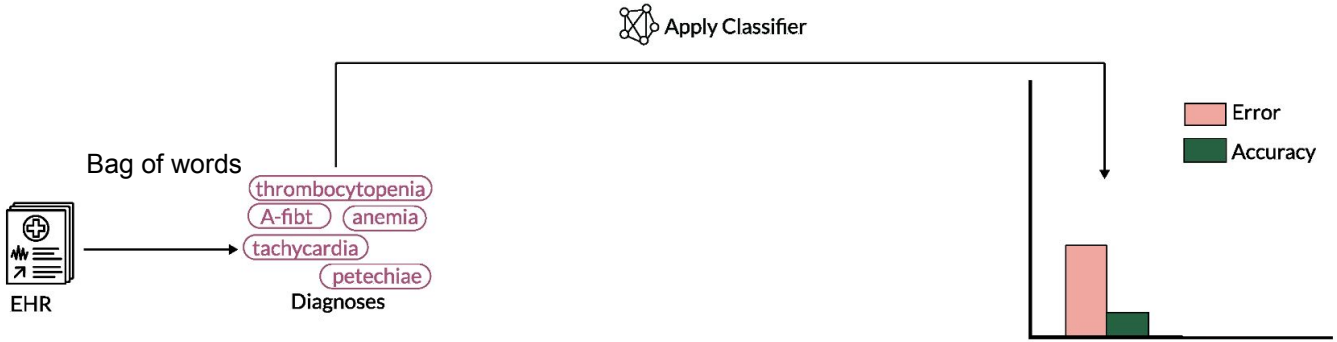
Legend: Color = disease, Shape = dataset origin
Banerjee et al, Nature Methods, 2023

Combine multiple datasets to make one big dataset



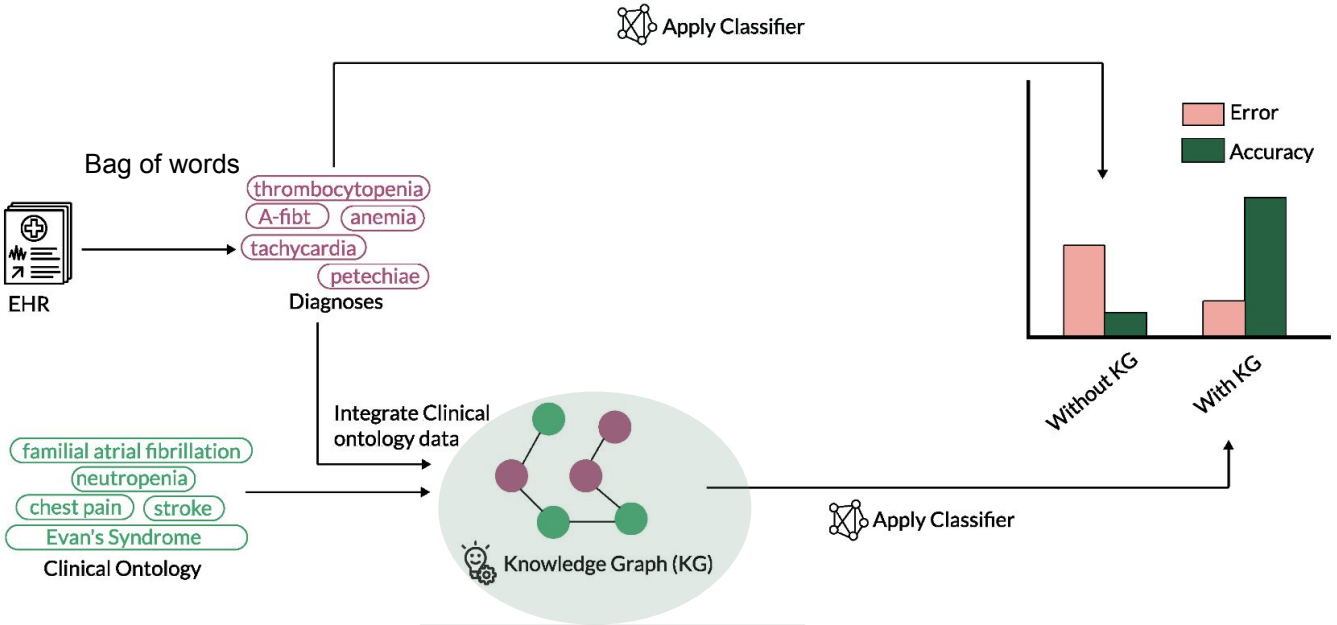
Legend: Color = disease, Shape = dataset origin
Banerjee et al, Nature Methods, 2023

Augment a dataset using prior knowledge



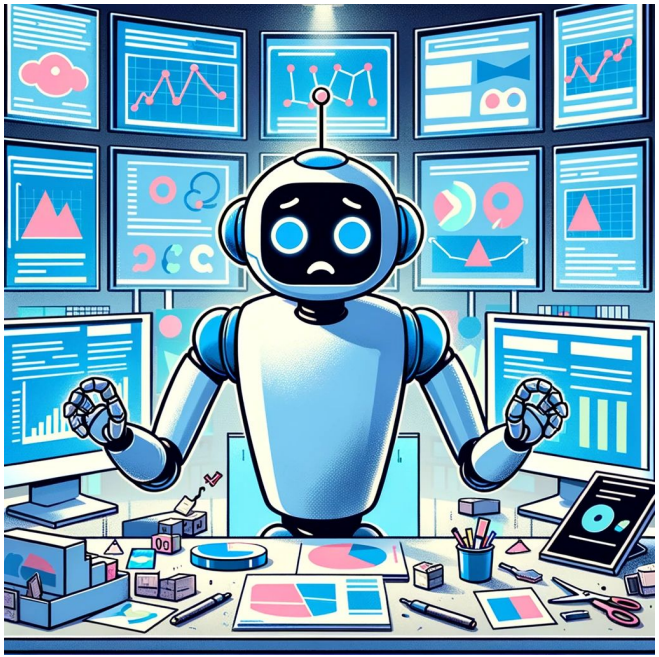
Li et al, 2019, BMC Med Inf and Decision Mak

Augment a dataset using prior knowledge



Li et al, 2019, BMC Med Inf and Decision Mak

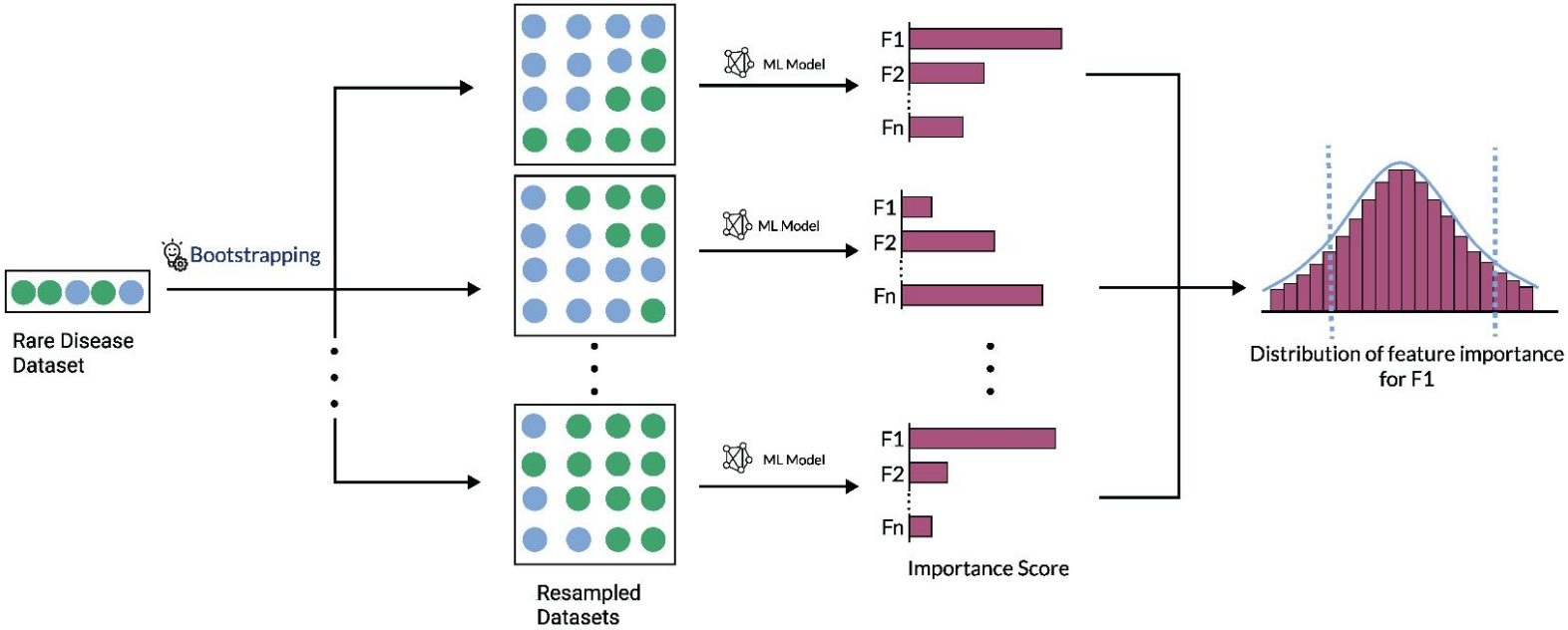
Methodological challenges for ML in rare disease



Images courtesy: DALL-E3

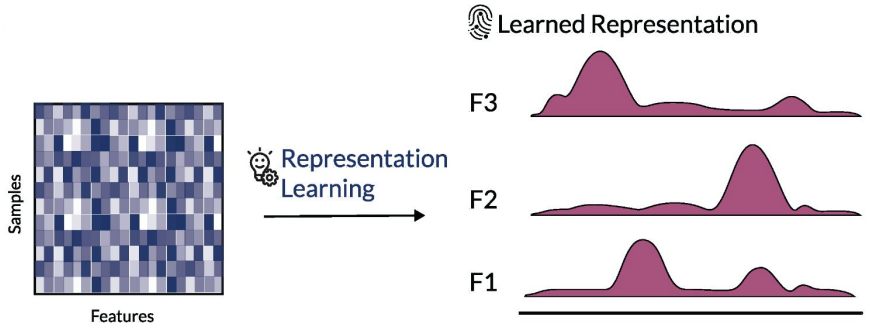
Artificially increase the exposure of the model to rare samples

Resampling with replacement



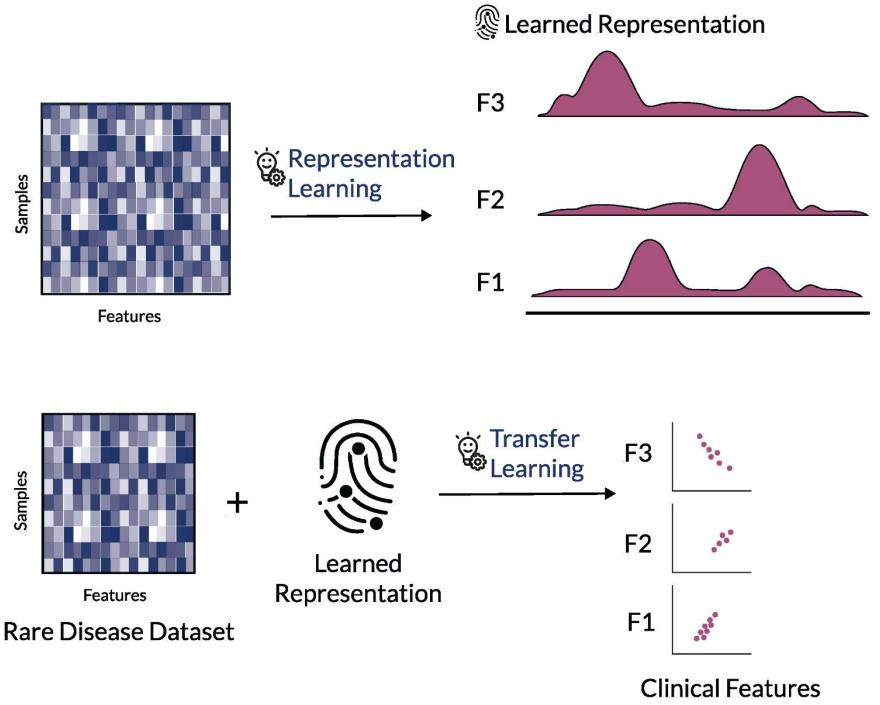
Learn from a larger dataset, then transfer to a small dataset

Transfer Learning

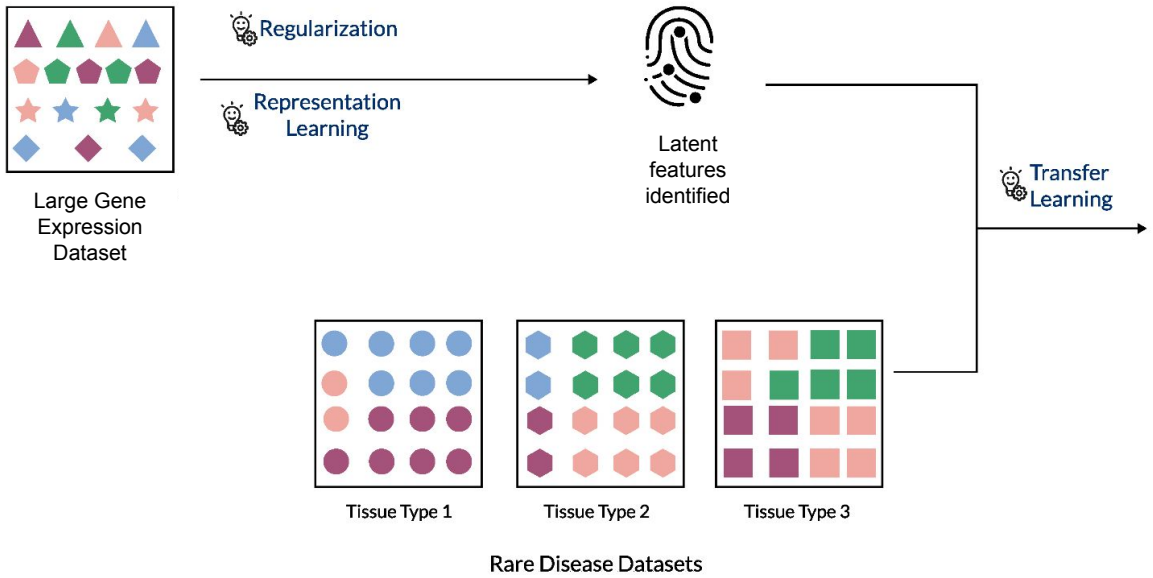


Learn from a larger dataset, then transfer to a small dataset

Transfer Learning



Combine multiple approaches



A single model to define the disease in multiple tissue types

Taroni et al, 2019, Cell Systems

Main takeaways:

Careful inspection of biomedical articles applying AI/ML models to rare disease :

- ➢ Were the datasets constructed appropriately?
- ➢ Were multiple approaches taken to build appropriate models for the task at hand?
- ➢ Were the methods statistically rigorous?

How can you help

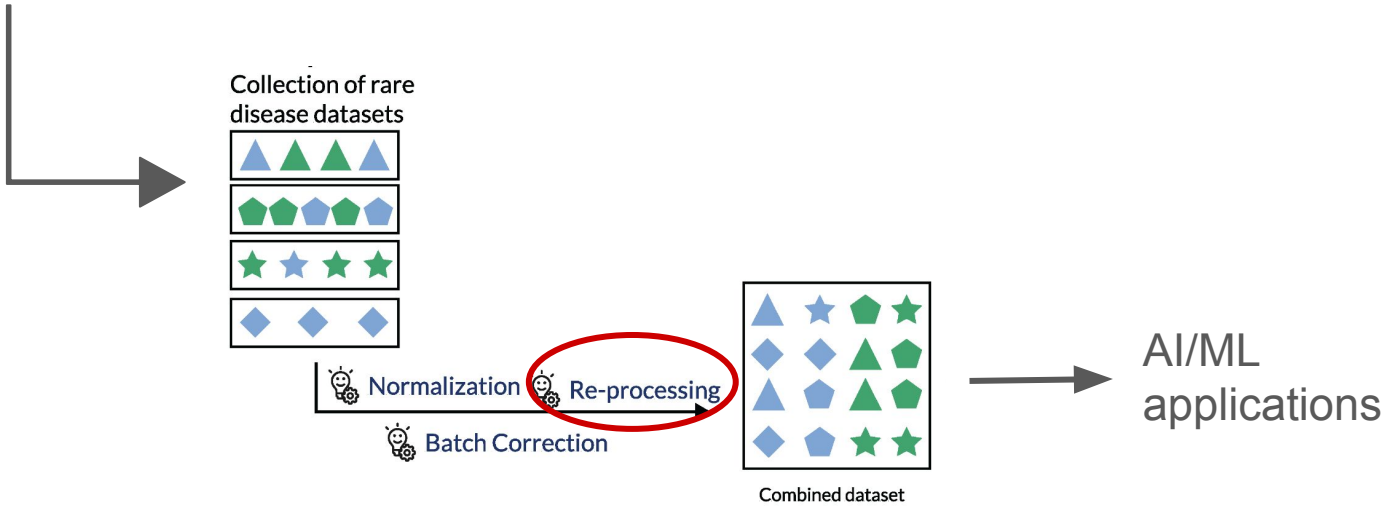
Help us get the word out :

- Need well-described data that were generated with ML applications in mind

Making rare disease datasets openly available

NF DATA PORTAL

Powered by Sage Bionetworks



How can you help

Help us get the word out :

- Need well-described data that were generated with ML applications in mind
- Need for development of methods that are tolerant of small sample sizes

Even bigger implication

Rare

Even bigger implication

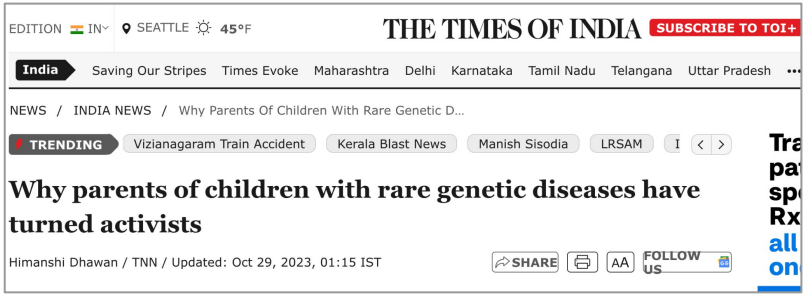
Rare

One among many

Even bigger implication

One among many = Personalized medicine

How can AI or ML help people living with rare diseases?



Images courtesy: DALL-E3

How can AI or ML help caregivers ?

- Caregivers turned patient-advocates
- Primary care physicians who may not be disease experts

Images courtesy: DALL-E3

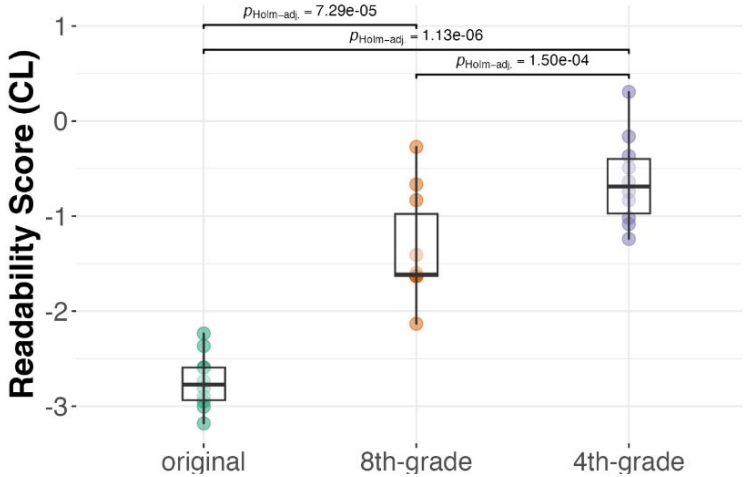
Enhance accessibility to rare disease research

- Patients and caregivers generally have to advocate for themselves
- Need to keep up with research in their disease
- Patients and caregivers are not used to scientific jargon

Images courtesy: DALL-E3

Enhance accessibility to rare disease research

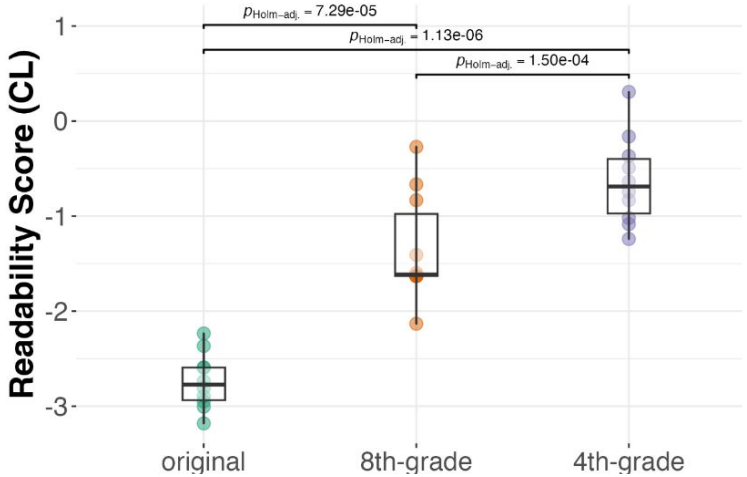
Large Language Models can help simplify research

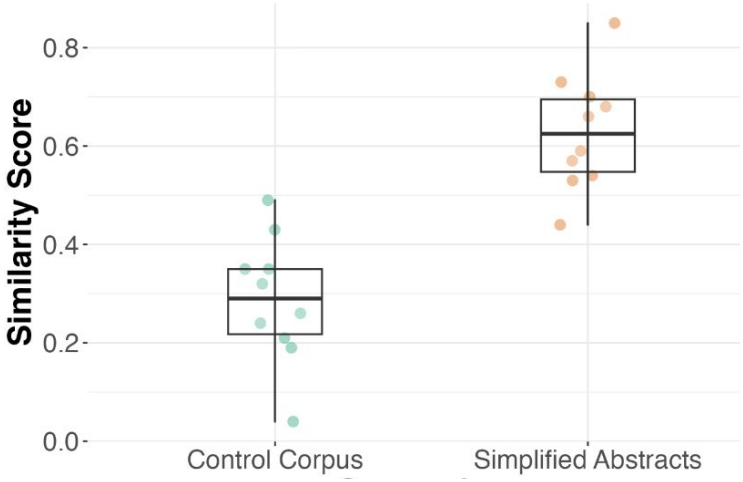


Robert Allaway

Enhance accessibility to rare disease research

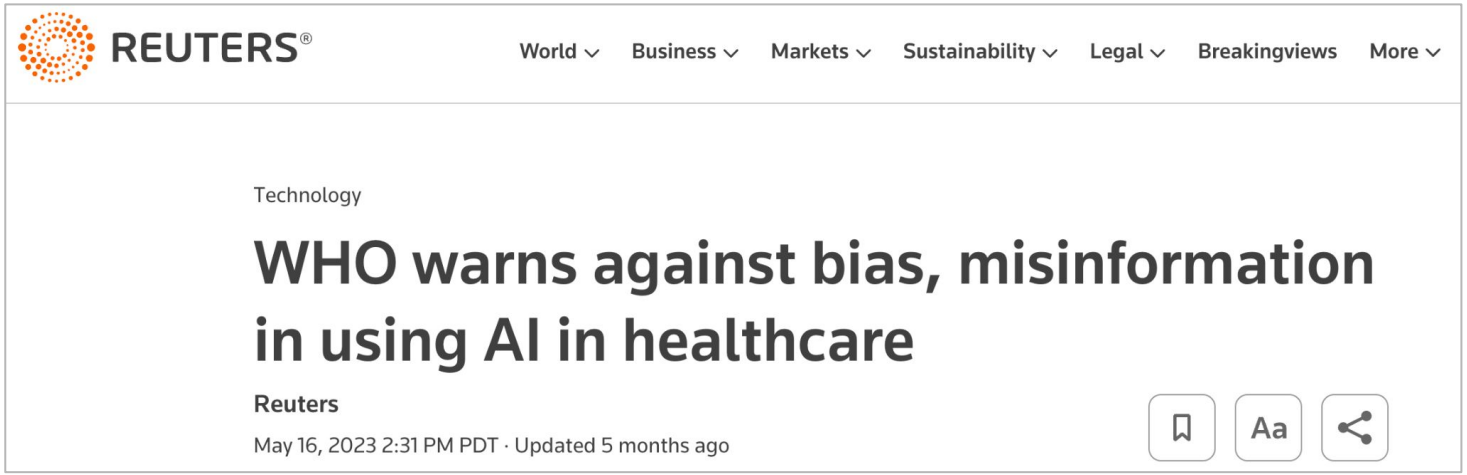
Large Language Models can help simplify research but preserve meaning





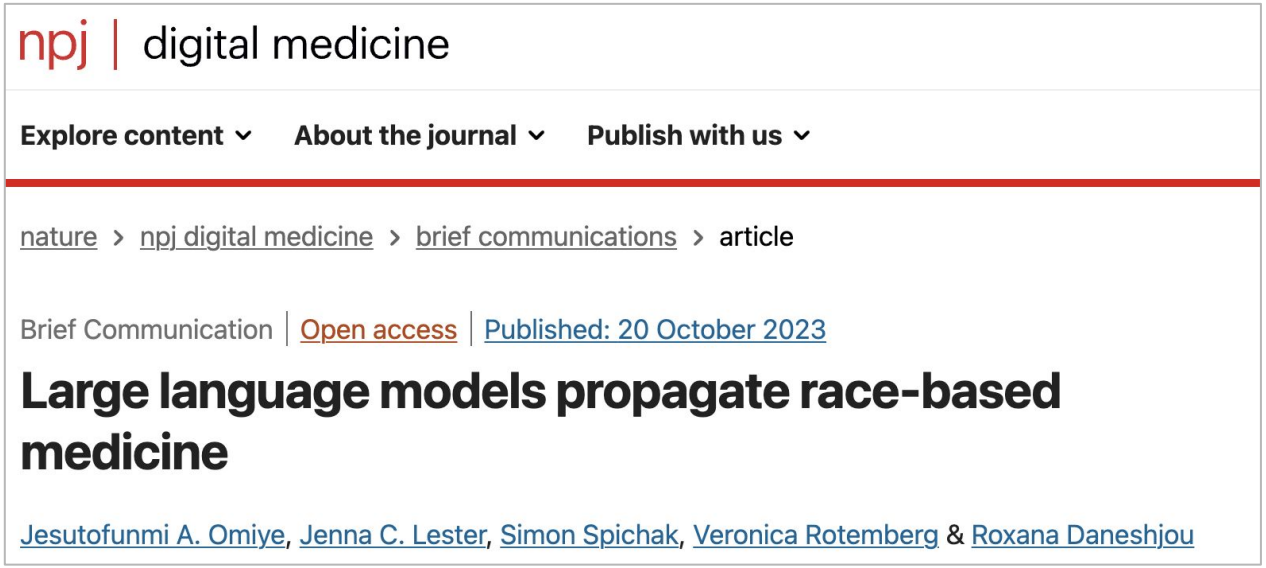
How can you help

- Vigilance towards incorrect information from LLMs



How can you help

- Vigilance towards inherent biases of LLMs and how they affect patients



npj | digital medicine

Explore content ⌄ About the journal ⌄ Publish with us ⌄

nature › npj digital medicine › brief communications › article

Brief Communication | Open access | Published: 20 October 2023

Large language models propagate race-based medicine

Jesutofunmi A. Omiye, Jenna C. Lester, Simon Spichak, Veronica Rotemberg & Roxana Daneshjou

An exciting time ahead!

Images courtesy: DALL-E3

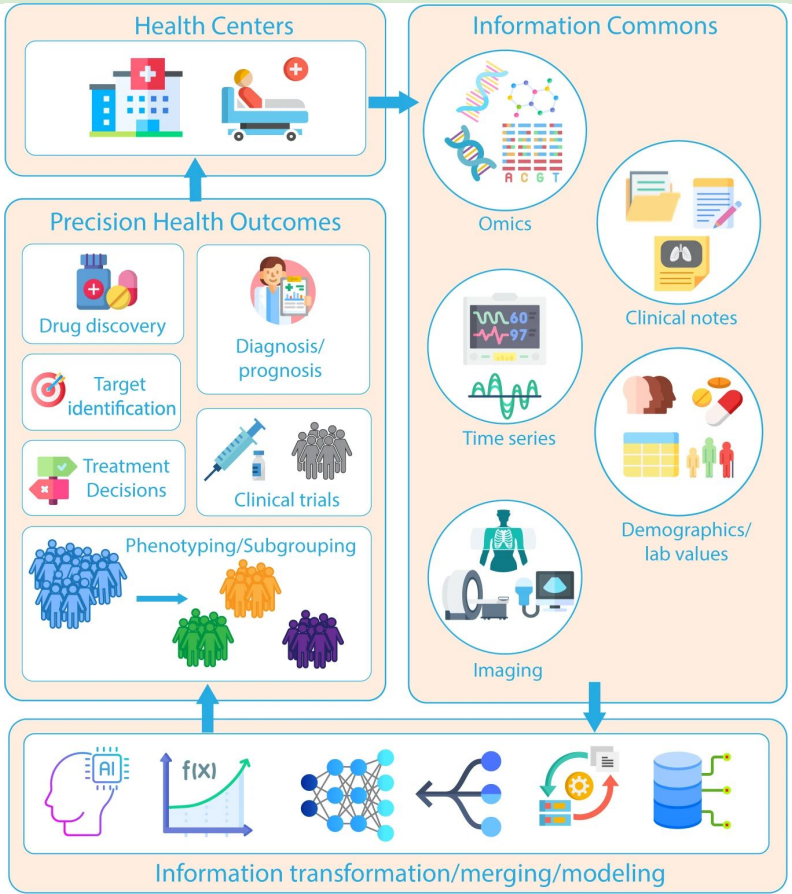
Looking to the future:

Multimodal data from patients

- Improved capabilities of generating rich datasets from rare disease patients

➢ Promises:

- Provide complete picture of disease



Kline et al, NPJ DiMe, 2023

Looking to the future :

Digital twins in rare disease

- Digital twin technology
  - Digital twins help test technical advancements before release to general public

McKinsey & Company

Industries | Capabilities | Featured Insights | Locations | Careers | About Us | McKinsey Blog

What is digital-twin technology?

July 12, 2023 | Article

Looking to the future :

Digital twins in rare disease

- Promises:
  - Disease modeling using multi-modal data
  - Protection due to advances in differential privacy and privacy preserving record linkage
  - Help in optimizing clinical trial design and pre-clinical research

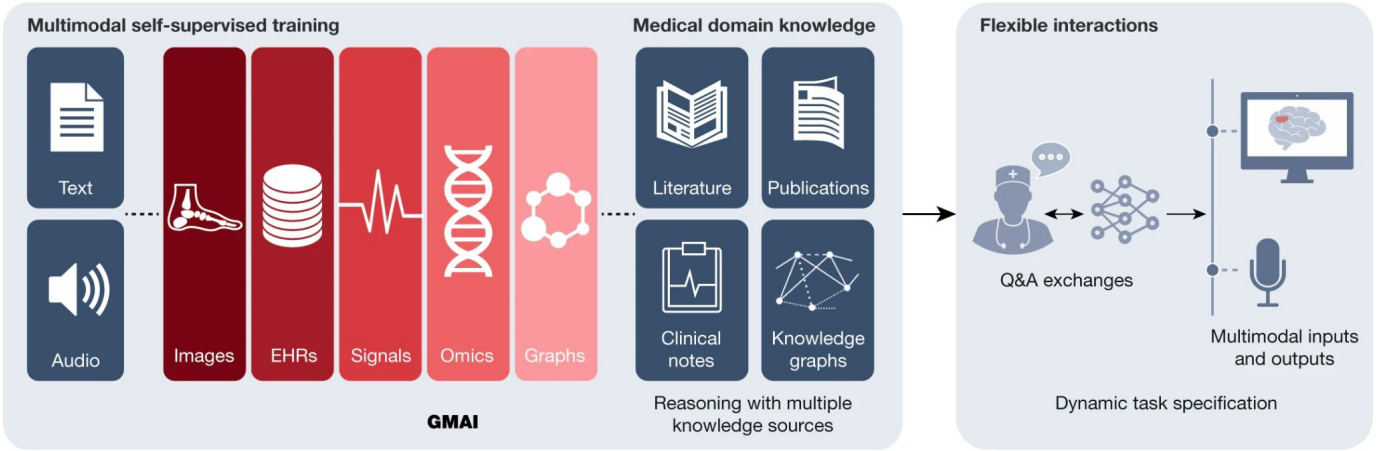
- Challenges to overcome :
  - Privacy
  - Ethical implementation
  - Reliability

Acosta et al, Nature Medicine, 2022
Images courtesy: DALL-E3

Looking to the future :

Foundation Models in Medicine

- Promises:
  - Flexible interactions
  - Adaptability

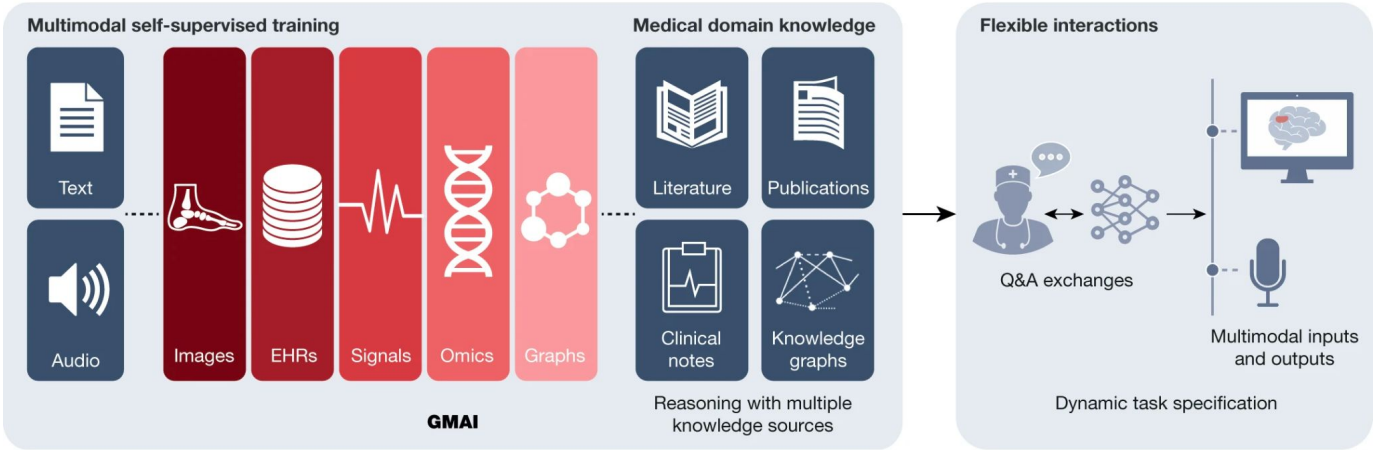


Moor et al, Nature, 2023

Looking to the future :

Foundation Models in Medicine

- Challenges to overcome:
  - Reliability
  - Biases
  - Privacy preservation



Moor et al, Nature, 2023

How can you help

Help us get the word out :

- Promote generation of well-described data from underrepresented populations to train the next generation of models
- Promote vigilance towards protection of privacy in these sophisticated AI applications
- Promote fact-checking of AI applications to ensure reliability of the models

Thank You

All my colleagues at Sage Bionetworks

Collaborators:

- *Casey S Greene (University of Colorado, Anschutz)*
- **Jaclyn P Taroni (Alex's Lemonade Stand Foundation)*
- **Robert J Allaway (Sage Bionetworks)*
- *Deepa V Prasad (Alex's Lemonade Stand Foundation)*
- *Justin Guinney (Tempus Labs)*

Support:

Thank You