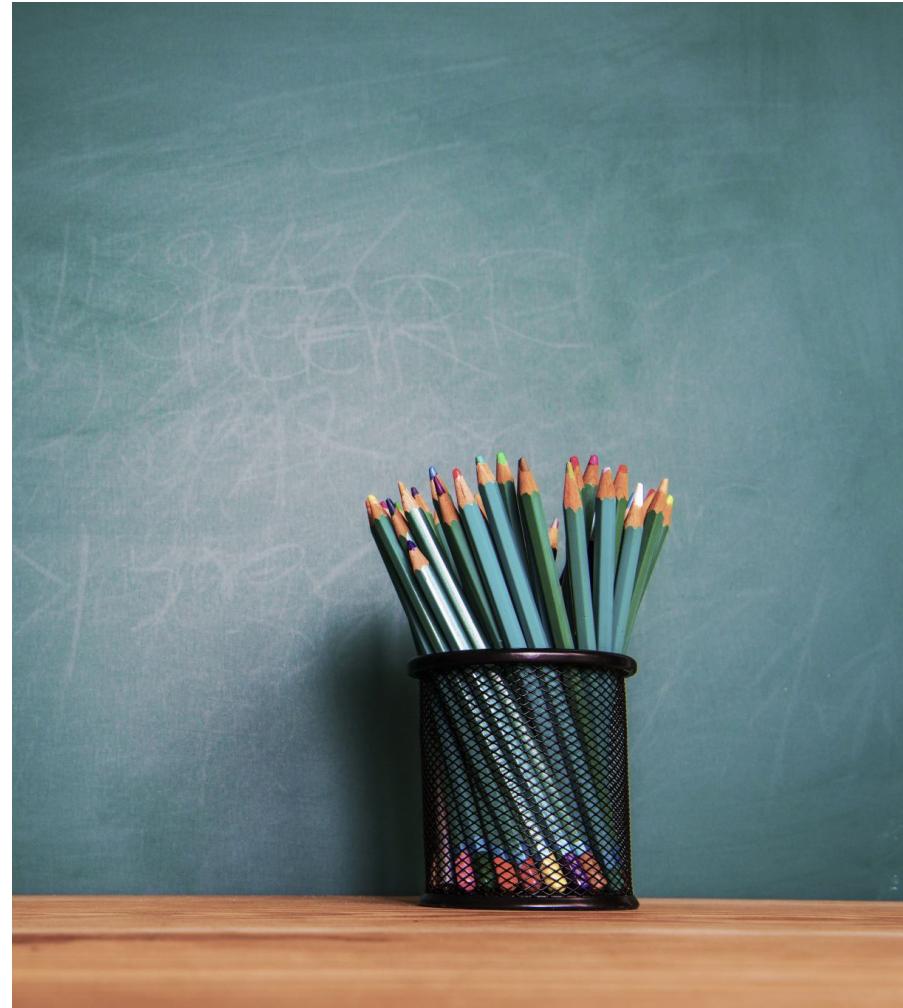
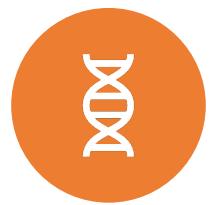


Patient and public involvement in Artificial Intelligence

Soumya Banerjee
University of Cambridge



Personal history



WORKED IN
COMPUTATIONAL
BIOLOGY



HAVE A PHD IN
COMPUTER
SCIENCE



WORKED IN CELL
BIOLOGY, CLINICAL
INFORMATICS, ETC.



ALSO WORKED IN
SUPPLY CHAINS
(WITH INDUSTRY)



DRUG
DEVELOPMENT
(WITH INDUSTRY)

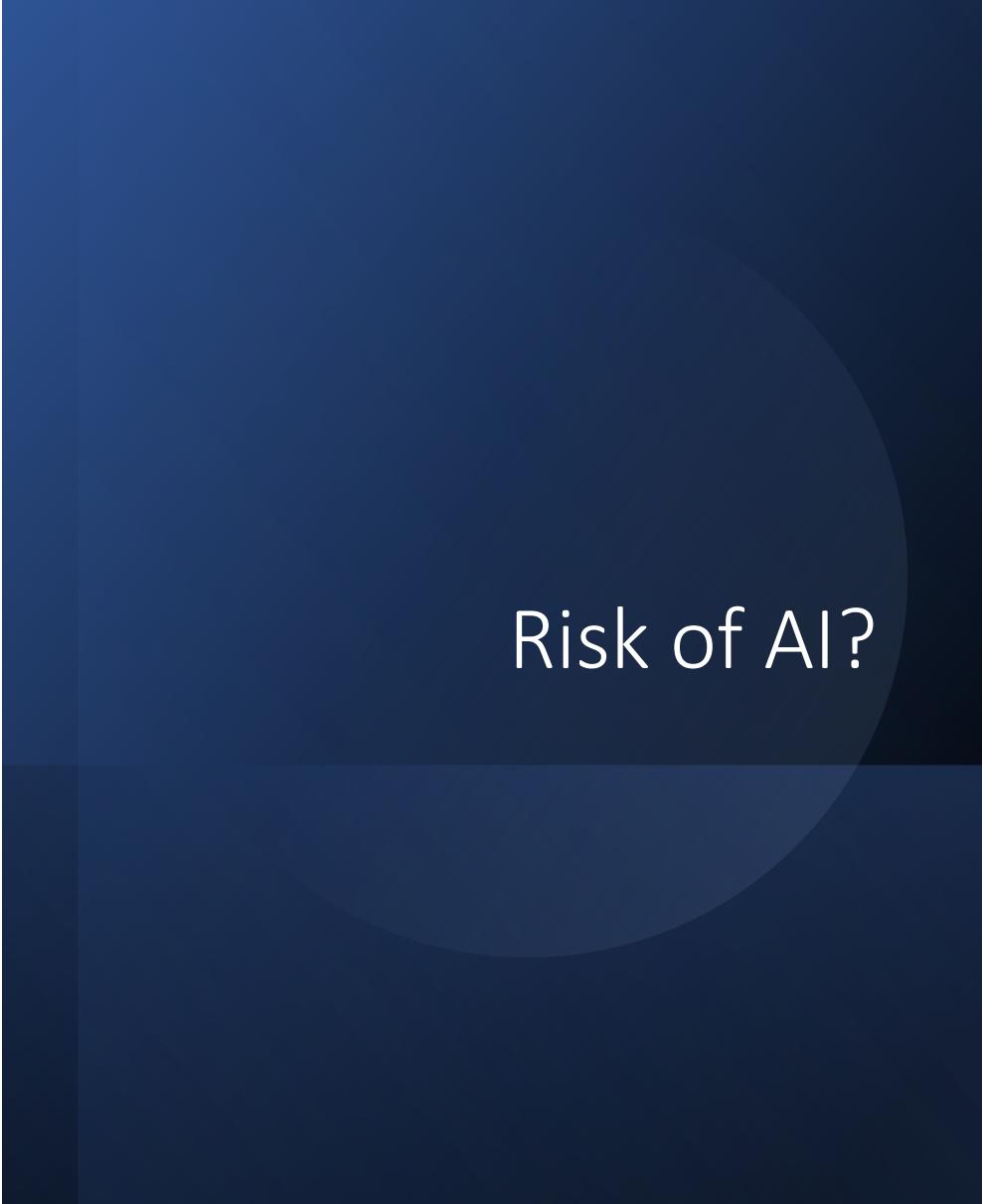
Background



The general public is disconnected
from AI and there is mistrust



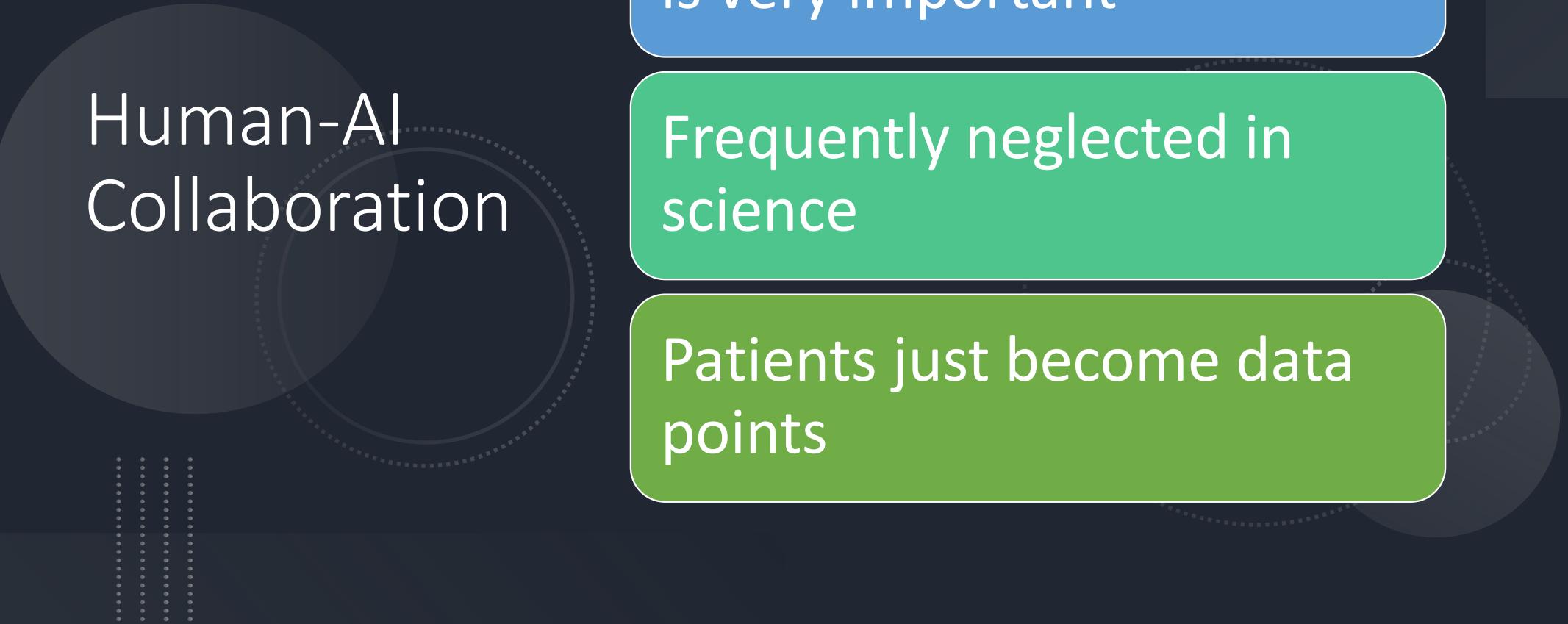
We do not trust what we do not
understand



Risk of AI?

- Critical stakeholders and end-users of AI or those who are impacted by AI are not involved in
 - *Understanding AI models*
 - *AI model building*

Human-AI Collaboration

The background features three large, semi-transparent circles in dark grey, black, and white, which overlap each other. A series of vertical dotted lines is positioned to the left of the text blocks.

The role of lived experience
is very important

Frequently neglected in
science

Patients just become data
points

One fine day ...



Patient
comes up
with
research
question!

Patient came up with hypothesis

Value of lived experience

Researchers (like me) frequently
do not know what it *feels* like to
have a disease

Steps

Recruiting patients
and forming a RAG

Problem
formulation and
hypothesis
generation

Building trust in
data storage
infrastructure

Addressing
concerns about big
data

Discrimination and
bias in AI

Debunking myths
about AI

Understanding a
simplified model

Understanding AI
models more
broadly

Designing
interfaces with
patients

The case of lithium

Data from hospital electronic healthcare records

EPIC (laboratory tests, diagnosis codes prescriptions)

Meditech (laboratory tests)

The case of lithium

$$e_{GFR} = e_0 + b_{off}t_{off} + b_{on}t_{on} + (1|pid + t_{off} + t_{on}) \quad (\text{Equation 1})$$

where e_{GFR} is the estimated glomerular filtration rate (eGFR) and is calculated from creatinine, age, gender, and ethnicity (data available from hospital electronic healthcare records system) using the CKD-EPI formula.³ pid is the unique patient identification number in the electronic healthcare record system. b_{on} is the rate at which eGFR declines when a patient is on lithium. t_{off} is the cumulative time spent off lithium, and t_{on} is the cumulative time spent on lithium. b_{off} is the rate at which eGFR is declining for patients off lithium, and b_{on} is the rate at which eGFR is declining for patients on lithium. e_0 , b_{on} , b_{off} , t_{on} , and t_{off} are parameters that are estimated from the data.

The case of lithium

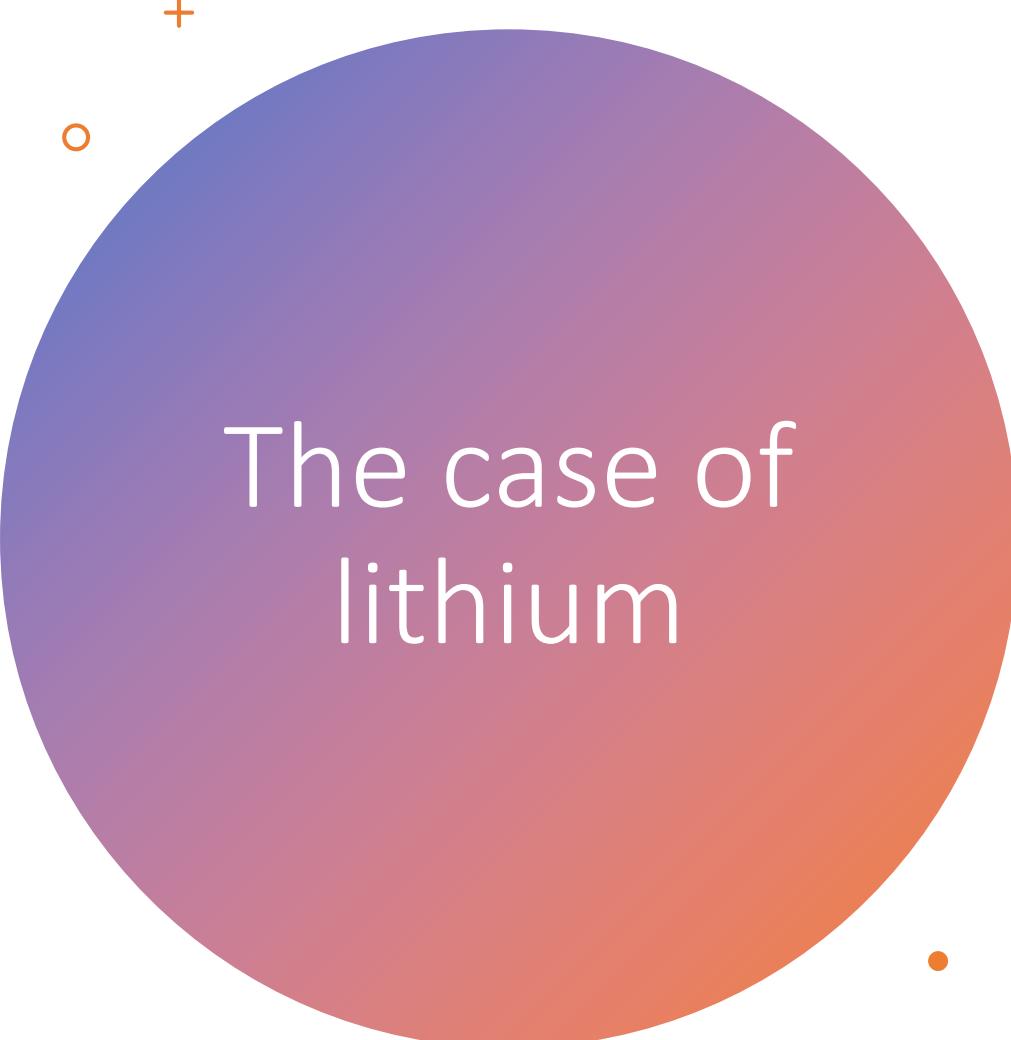
- We also took the time to explain how to build statistical models.
- For example, we tried other, simpler formulations before we arrived at the final model. This showed patients how researchers always incrementally build more complex models

$$e_{GFR} = e_0 + b_{off}t_{off} + b_{on}t_{on} + (1|pid) \quad (\text{Equation 2})$$

We explained these models using an example of a simpler linear model:

$$y = a \cdot x + b \quad (\text{Equation 3})$$

This is a linear model where the value of x is used to predict y (say eGFR). a and b are the parameters of the model, and these can be estimated. We explained that estimating means determining the values of a and b from data. Once we explained the concepts of a linear model, we progressed to more advanced concepts like confidence intervals.



The case of lithium

- However, using these data on a few thousand patients, the results were inconclusive.
- This motivated the need to go back to the RAG and explain the need for more data.
- We took feedback from patients as to whether we should apply for access to more data.
- We also built a tool that explains how, in some cases, having more data can help in estimating parameters of statistical models
- This process of performing research and getting inconclusive results also showed patients how research always takes time and can lead to unexpected roadblocks

Data Privacy

- How can big data help?
- Privacy and security explained
- Difficult to get access to data even for researchers
 - Saga of data access

Banerjee et al., *Patterns*, 2022



Bias in Algorithms

- Bias in face detection

Banerjee et al., *Patterns*, 2022



Tools for involvement and outreach



AI MODELS THAT CAN BE TRAINED, RUN, AND VISUALIZED IN THE WEB BROWSER, LIKE THE TEACHABLE MACHINE



AN AI MODEL THAT RUNS IN THE WEB BROWSER AND USES A WEBCAM TO DETECT FACIAL EXPRESSIONS



A SET OF VIDEOS THAT DEMONSTRATE WHAT AI CAN AND CANNOT DO



A RESOURCE OF MYTHS ABOUT AI



https://github.com/neelsoumya/ai_outreach

https://github.com/neelsoumya/ai_outreach

Communicating caveats

At this stage, we also communicated to the patients a number of caveats. Lithium is an effective medication for managing bipolar disorder and the chances of patients developing renal complications is quite small.

The benefit of discontinuing lithium should be carefully weighed against the risk of relapse of the psychiatric disorder, as has been documented in case studies and suggested in meta-analysis studies.

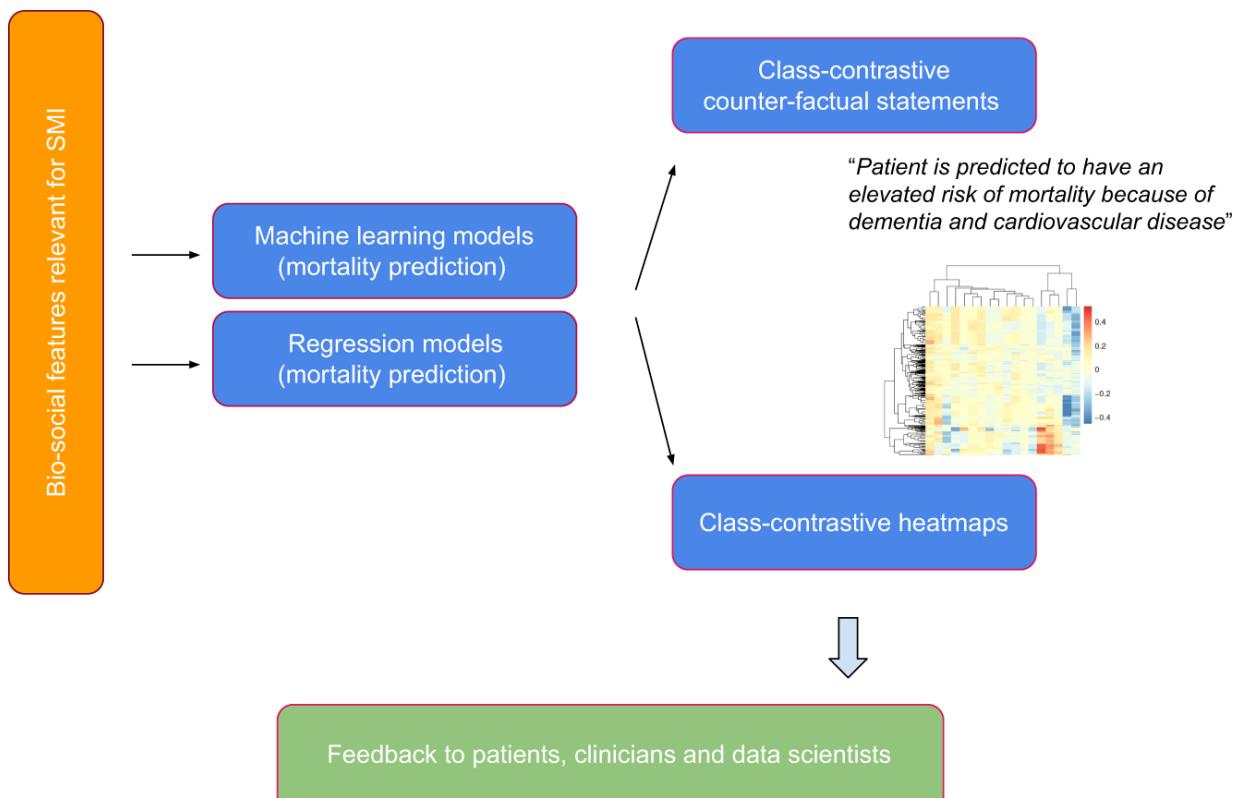
Our work may lead to randomized controlled trials to test the hypothesis that discontinuing lithium may help recover kidney function in patients with bipolar disorder.

The case of developing an xAI technique

- Patients with schizophrenia
- xAI technique
- Visual and textual explanation
- Workshop

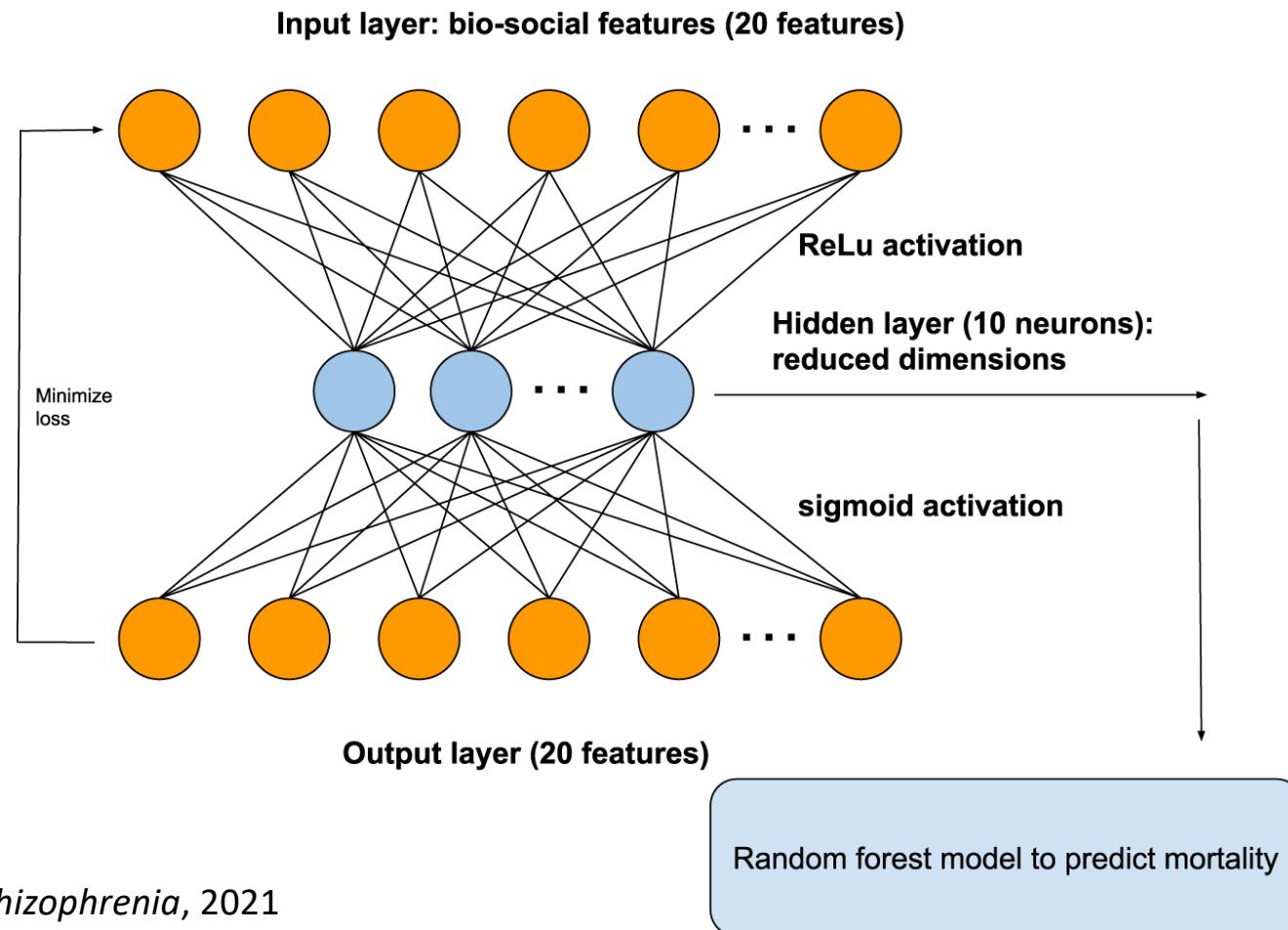


The case of developing an xAI technique

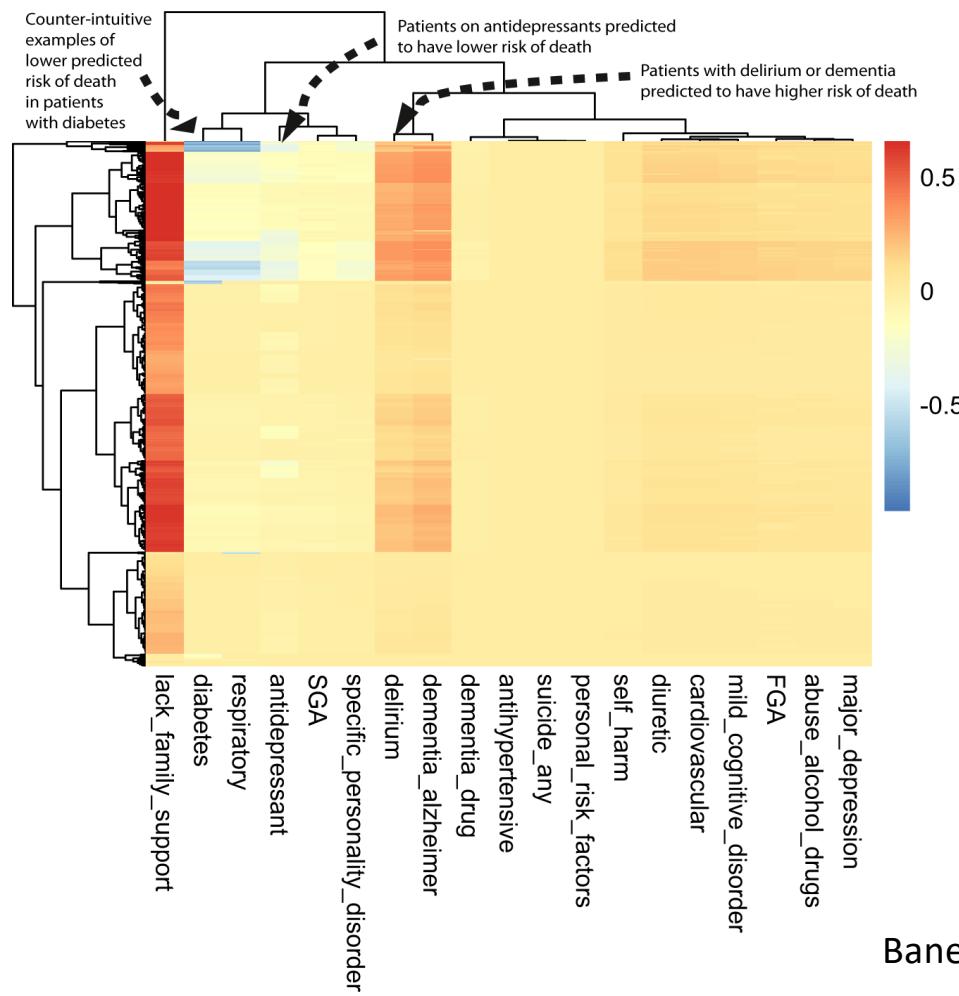


Banerjee et al., *NPJ Schizophrenia*, 2021

The case of developing an xAI technique



The case of developing an xAI technique



Banerjee et al., *NPJ Schizophrenia*, 2021

Patient perspective

Patients and carers have important research ideas about how best to improve quality of life, manage symptoms, offer existing treatments, or develop new interventions.

Often these ideas differ from those prioritized by academia or the pharmaceutical industry. Here is why two members got involved in research in our group:

"I decided to join the group to help make a difference using my experience with a mental health condition, in my case being diagnosed with bipolar in 2003. To offer and share ideas and tips in what helps me and also share my experience to help research in the future."

"I was excited when I saw the invitation to join this group. Using more extensive data can potentially answer many vital questions that an 8-week drug trial simply cannot. As a service user I was keen to see how we can be involved."

Limitations

Costly to do

Time-consuming

Researchers not incentivized (does not count towards REF or tenure)

Quantitative training for patients

Long-standing collaborations required and trust

- Takes time

Efficacy? How do we know if this *works*?

- Human factors study

Systemic issues



Researchers are not incentivized to do PPI or outreach



Structural inequalities and power hierarchies in academia



Risky to work in this area (especially for early career researchers)

Why inform the public?



Public needs to be more informed about AI

More science awareness needed

COVID vaccine hesitancy

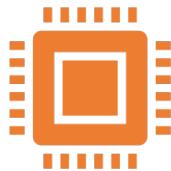
Fake news/misinformation



Resources

https://github.com/neelsoumya/ai_outreach

Why inform the public?



Public needs to be involved

Netherlands (benefits fraud detection system)

Fujitsu scandal in UK

Monitoring (and human rights violation) of Uyghurs in China



Citizen's jury



Involving the public and end-users

May lead to trustworthy AI systems



We do not
trust what
we do not
understand



Solutions

Allow the public to build models and develop their own understanding

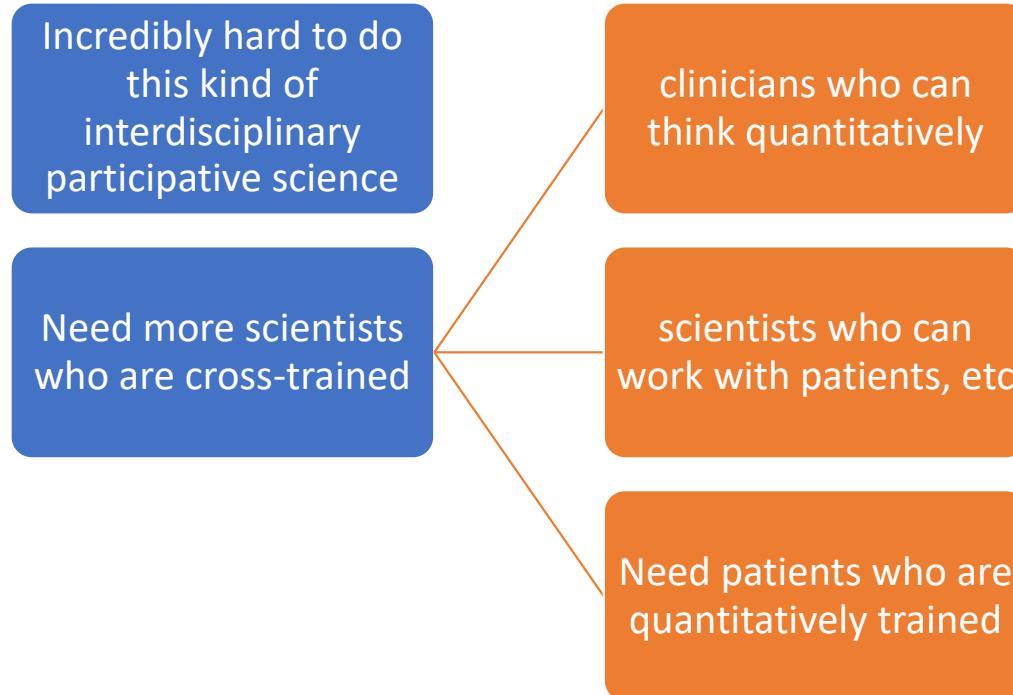
Power to the people

AI of the people, by the people, for the people

Citizen jury

Applicable to areas beyond healthcare

Solutions



The Power of Curiosity

- Let people/patients explore models on their own
- https://github.com/neelsoumya/ai_outreach

Summary

Outlined case studies and a methodology of how modern data science can be applied to healthcare where data scientists, clinicians, and patients work together.

The strategy of co-designing AI algorithms with patients is a balanced approach. This can help set more realistic expectations for all stakeholders since conventional narratives of AI revolve either around dystopia or limitless optimism.

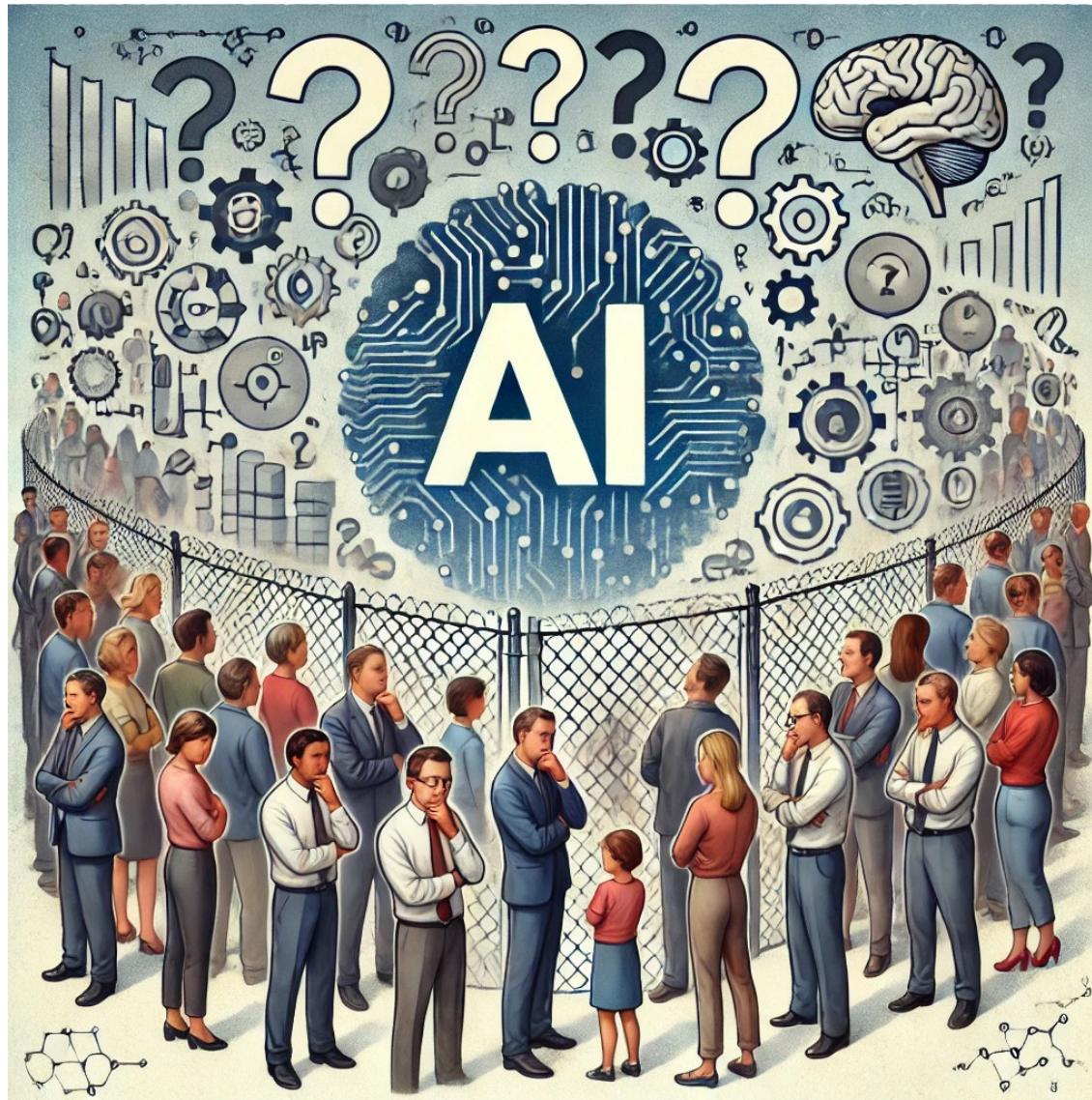
We hope that AI research in healthcare can be adopted faster if humans slowly build up trust in machines, over repeated and carefully calibrated interactions. May help build trustworthy AI.

Patients, data scientists, and healthcare workers can work together, thus benefiting patients.



Papers and Resources

- **Patient and public involvement to build trust in artificial intelligence: a framework, tools and case studies,** Soumya Banerjee, Phil Alsop, Linda Jones, Rudolf Cardinal, *Patterns* 3(6):100506, 2022
- **Involving patients in artificial intelligence research to build trustworthy systems,** Soumya Banerjee, Sarah Griffiths, *AI & Society*, 2023
- **A class-contrastive human-interpretable machine learning approach to predict mortality in severe mental illness,** Soumya Banerjee, Pietro Lio, Peter Jones, Rudolf Cardinal, *Nature Partner Journal Schizophrenia*, 7, 60, 2021
- https://github.com/neelsoumya/ai_outreach
- neel.soumya@gmail.com



Current Trends and Applications

- Deep learning
- Generative AI
- Applications in Healthcare, Finance, Autonomous Systems

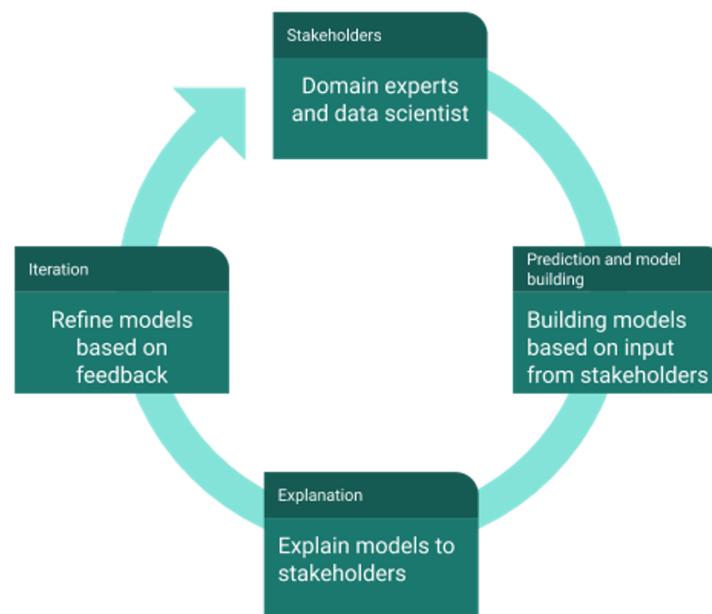
Risks of AI

- Ethical considerations
- Bias
- Data privacy
- Impact on jobs
- Systemic effects on our planet

Summary



Research and teaching philosophy



The Importance of Collaboration

Personal Experiences in AI

Importance of Asking Questions

Learning from Setbacks

Value of Teamwork and Networking

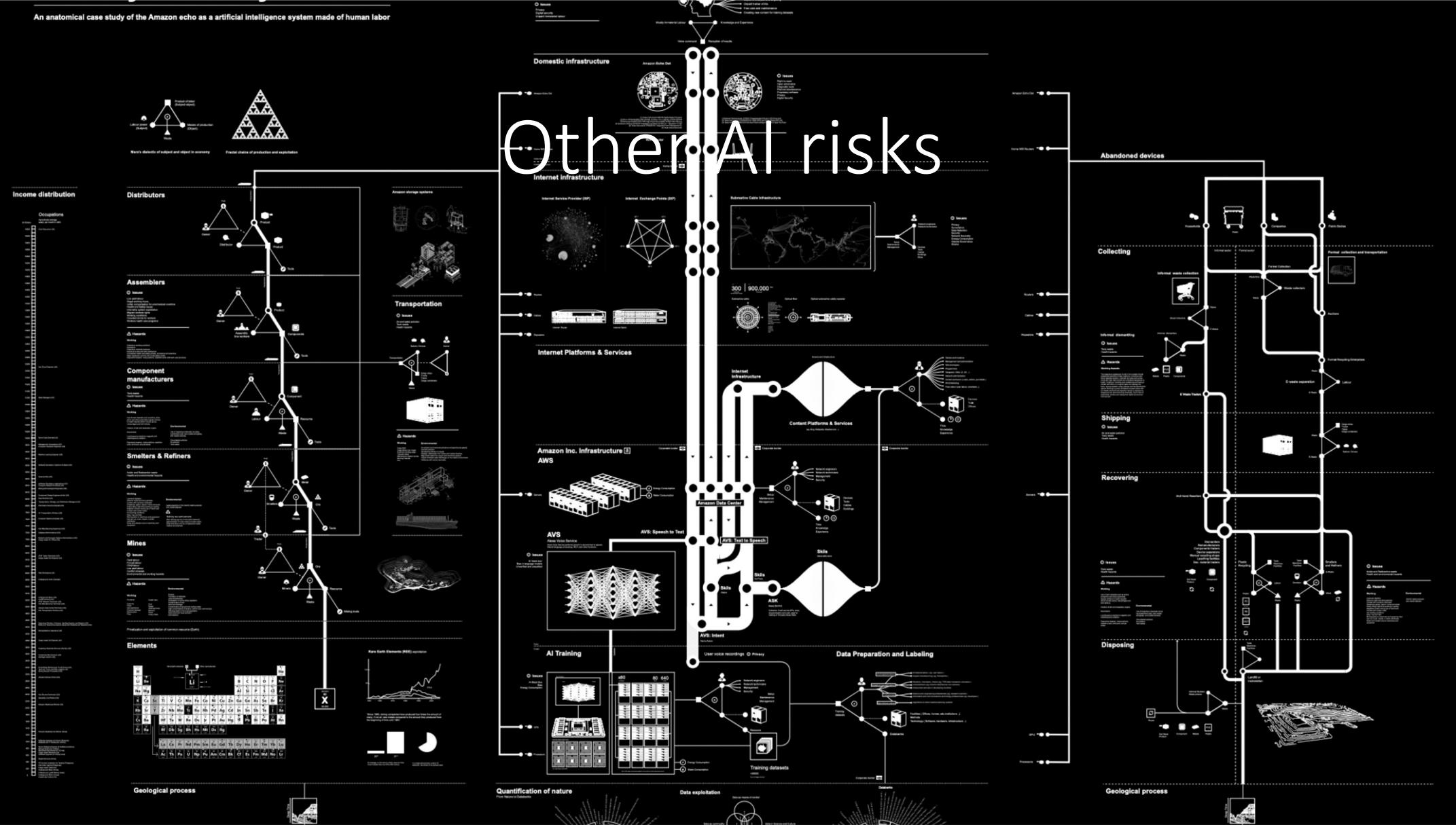
Balancing Theory and Practice

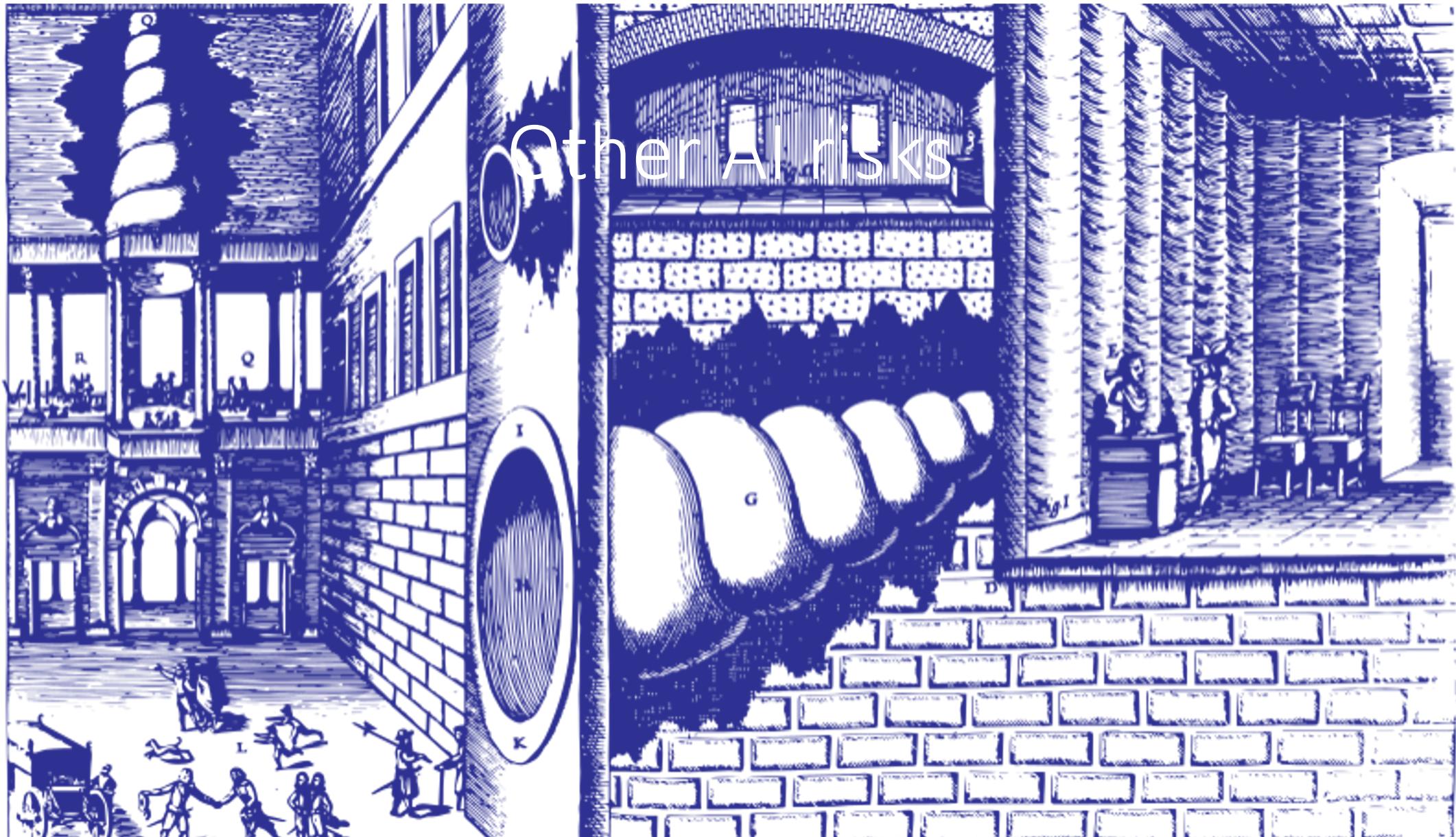
Combining Learning with Practical Experience

Looking Ahead

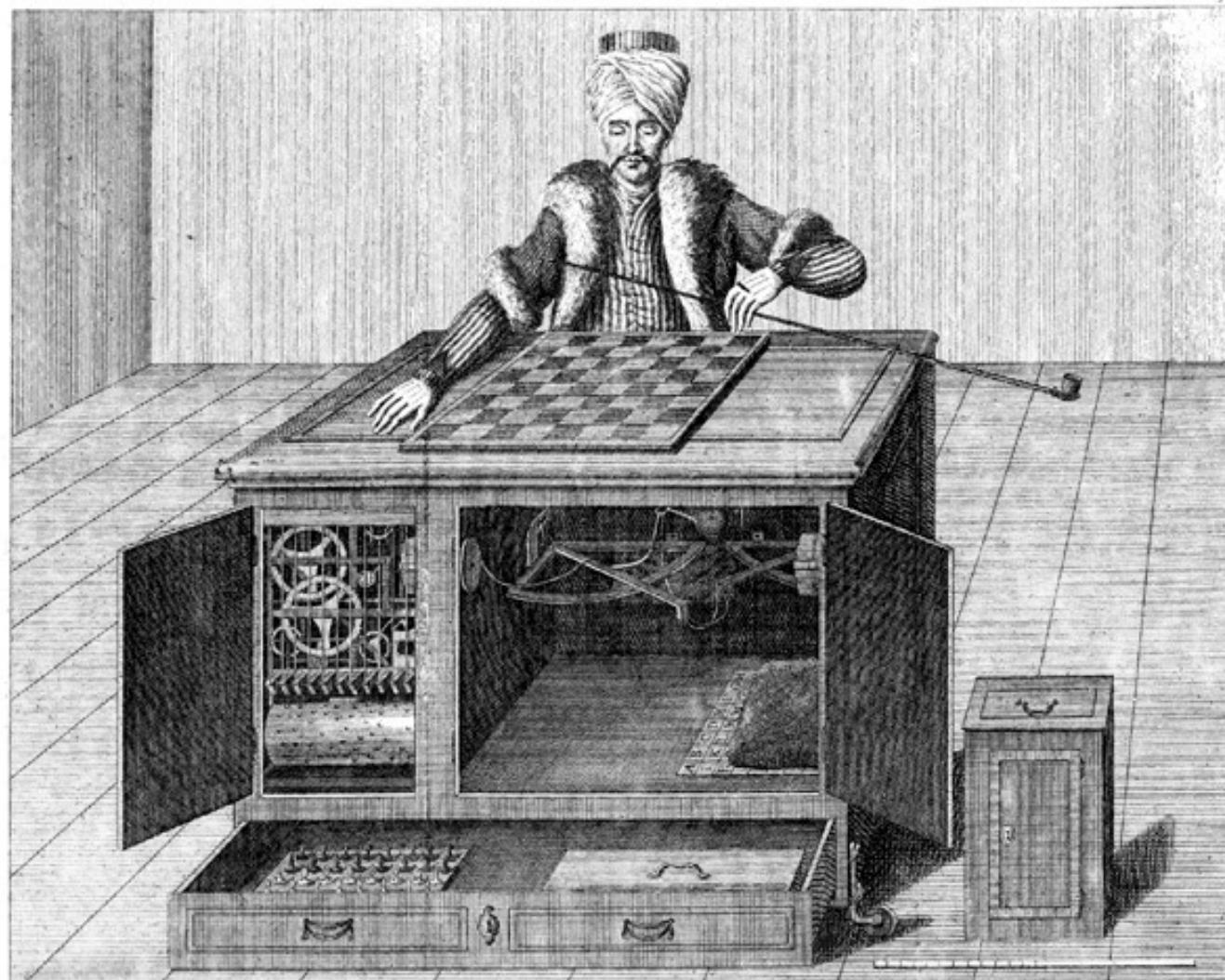
Impact of AI on Global Challenges

An anatomical case study of the Amazon echo as a artificial intelligence system made of human labor





Other AI risks

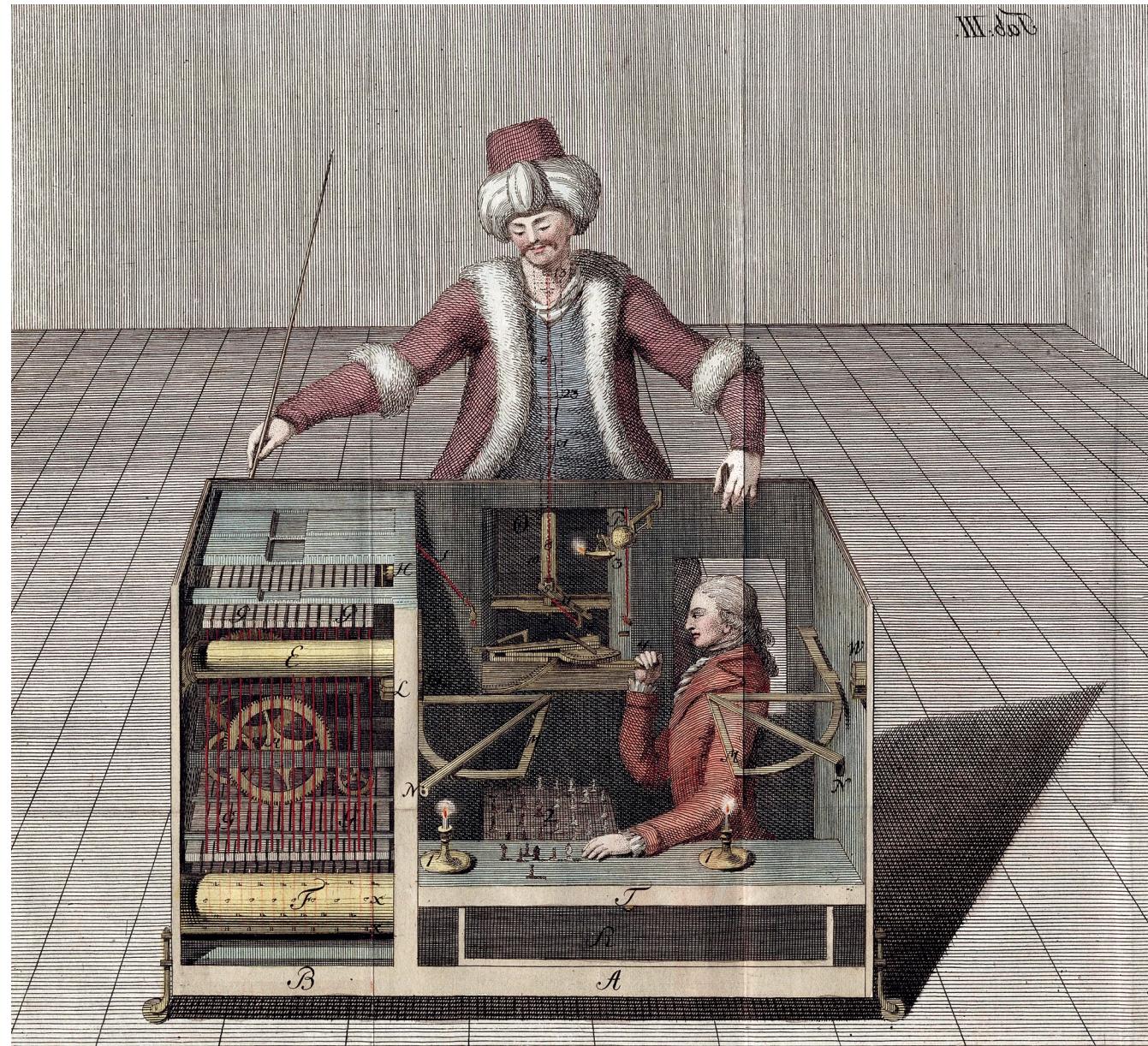


W. de Kempenaer del.

Chr. a Mechel excud. Basilew.

P.G. Pintz sc:

Der Schachspiele, wie er vor dem Spiel gezeigt wird, vornewm Le Jouer d'echec, tel qu'on le montre avant le jeu, par devant.



AI: An illusion?



The Digesting Duck automaton, created in the 18th century by Jacques de Vaucanson, was a marvel of mechanical ingenuity, designed to mimic the actions of a real duck, including eating, digesting, and excreting food. Similarly, ChatGPT, an advanced AI language model developed by OpenAI, is a modern marvel of artificial intelligence, designed to mimic human conversation and thought processes. Both the Digesting Duck and ChatGPT operate on the principle of simulating life-like behavior through complex internal mechanisms—whether mechanical gears or intricate algorithms. However, just as the duck's "digestive process" was ultimately a clever illusion, ChatGPT's understanding of language is an impressive simulation rather than true comprehension. Both serve as remarkable examples of their respective eras' technological advancements, pushing the boundaries of what machines can do to emulate natural processes.

Answer by ChatGPT

The first vending machine

The first vending machine





Any sufficiently advanced technology
is indistinguishable from magic

AI: Magic?

- Magic is done by magicians
- Magicians (AI scientists and big-tech) can get in a position of privilege and abuse
- Understanding AI and opening the black-box is important
- Common people also need to understand what is AI

Should not substitute people with AI

- Fujitsu scandal
- Prosecutions and Convictions: Based on the evidence provided by the faulty Horizon IT system, the Post Office prosecuted around 736 postmasters for theft, fraud, and false accounting. Many of these individuals were convicted, with some even serving prison sentences.
- For years, postmasters insisted that they had not committed any crimes and that the discrepancies were due to issues with the Horizon system.
- However, their concerns were dismissed by the Post Office, which relied on the Horizon system's data to assert their guilt.
- Takeaway: Decisions should be made by people, not machines



Big tech in AI: boon or curse?

- Big tech
- Creates jobs
- Too much power getting concentrated in the hands of a few companies
- We do not want a repeat of the East India company

Future Directions

- Explainable AI
- Human AI collaboration

My journey and personal experiences

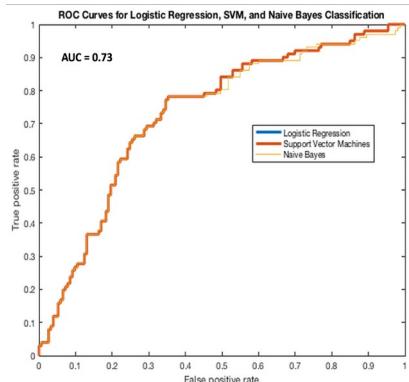
Clinical Informatics for Serious Illnesses

Clinical genomic data



Stratify patients

Anonymization and security procedures



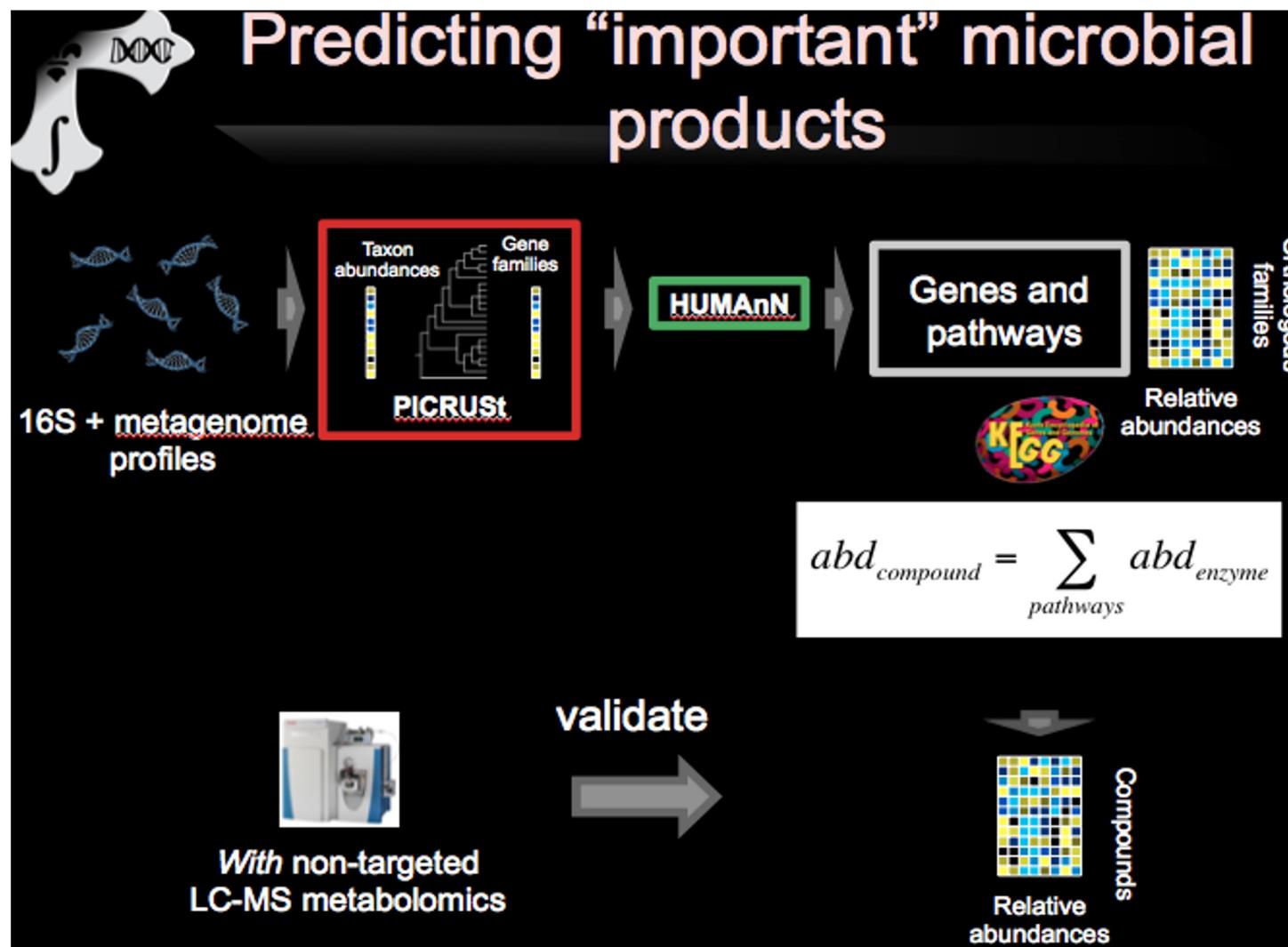
Machine Learning



Clinical metadata

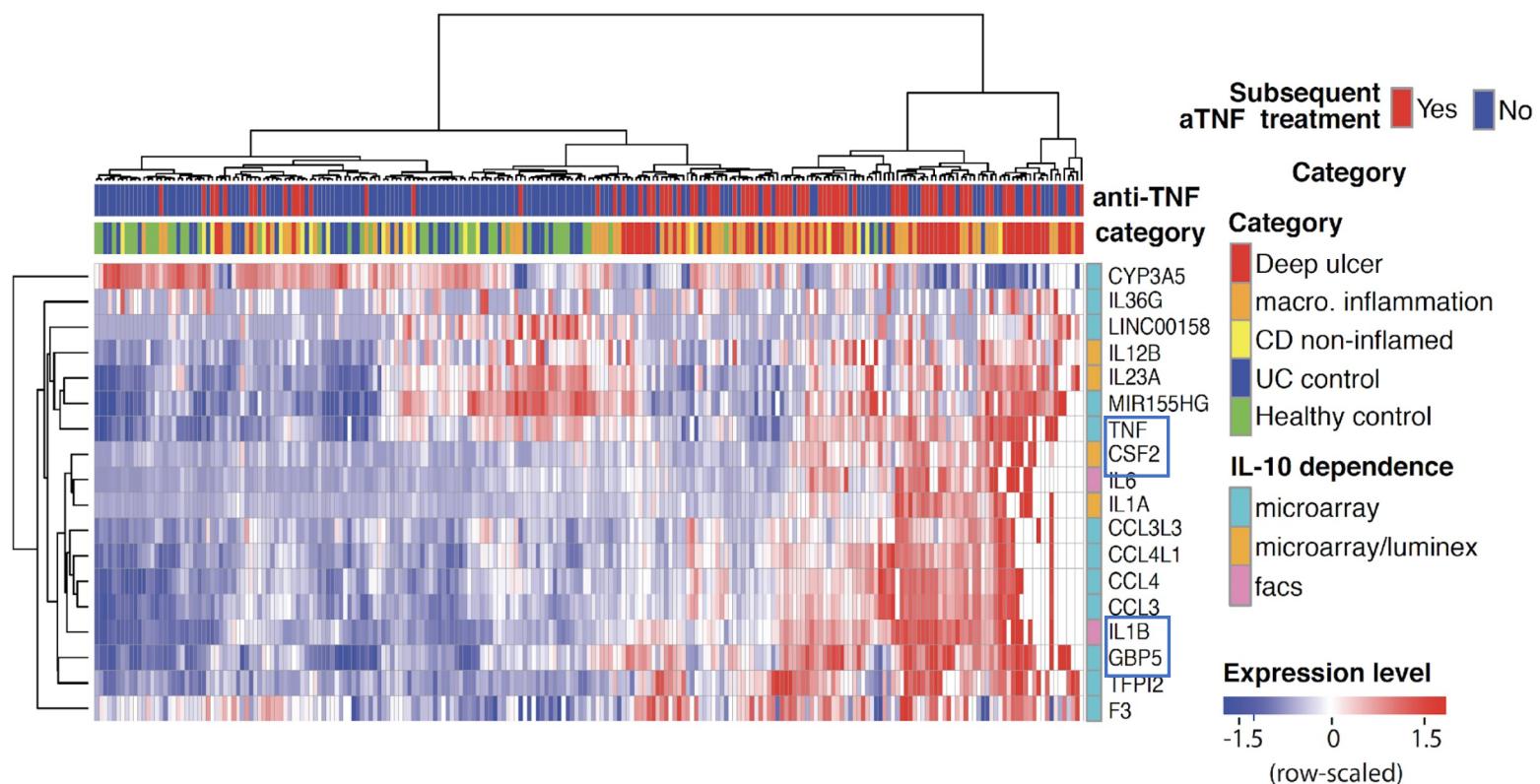


Input from domain experts (clinicians, statisticians, patients)



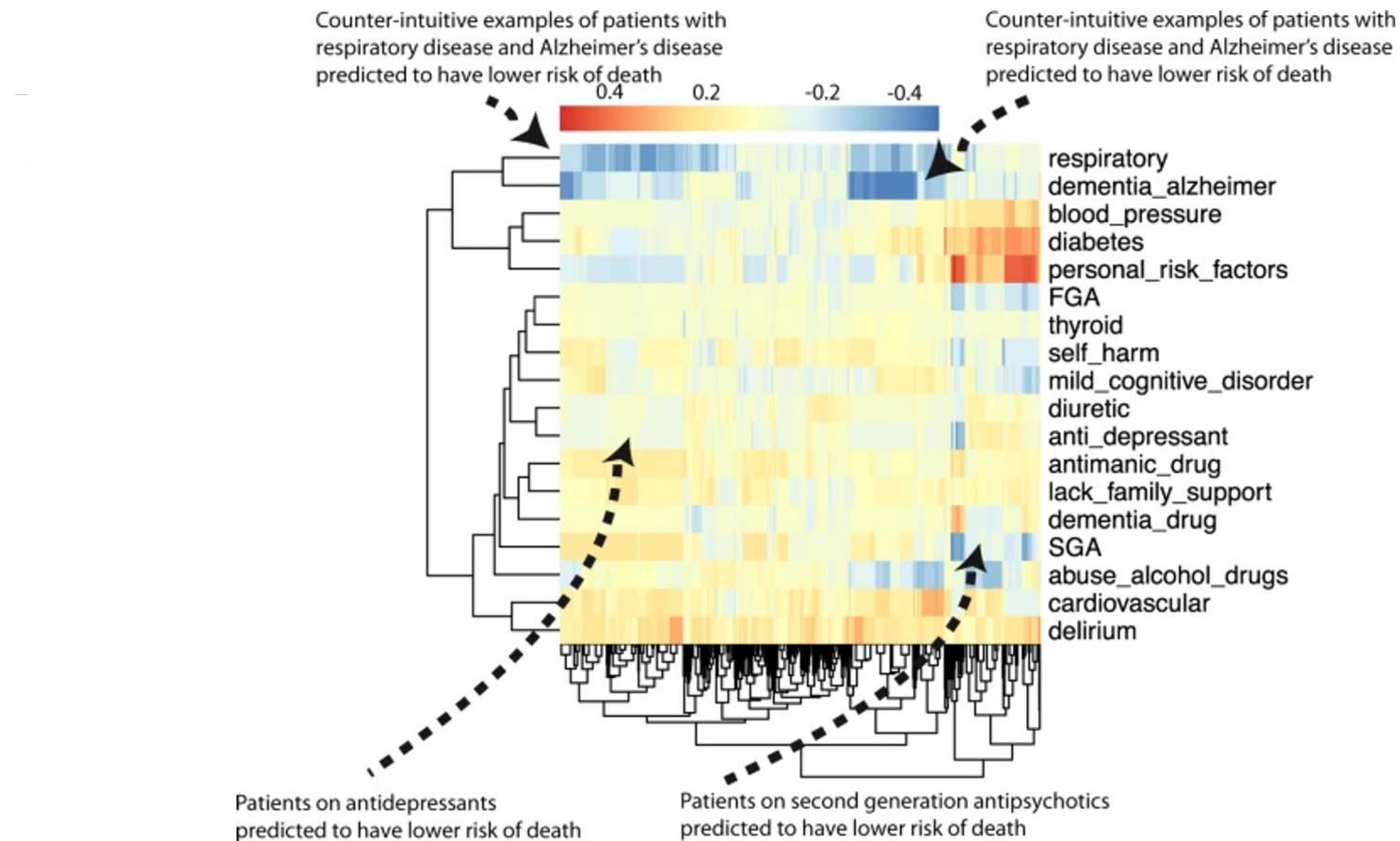
Nature Communications, 2019 (Nature publishing group), Impact factor = 14.9

Using machine learning algorithms to predict disease



Gut, 2020 (British Medical Journal publishing group),
Impact factor = 19

Interpretable machine learning



NPJ Schizophrenia, 2021 (Nature Partner Journal), Impact factor 6.6

Using machine learning in supply chains

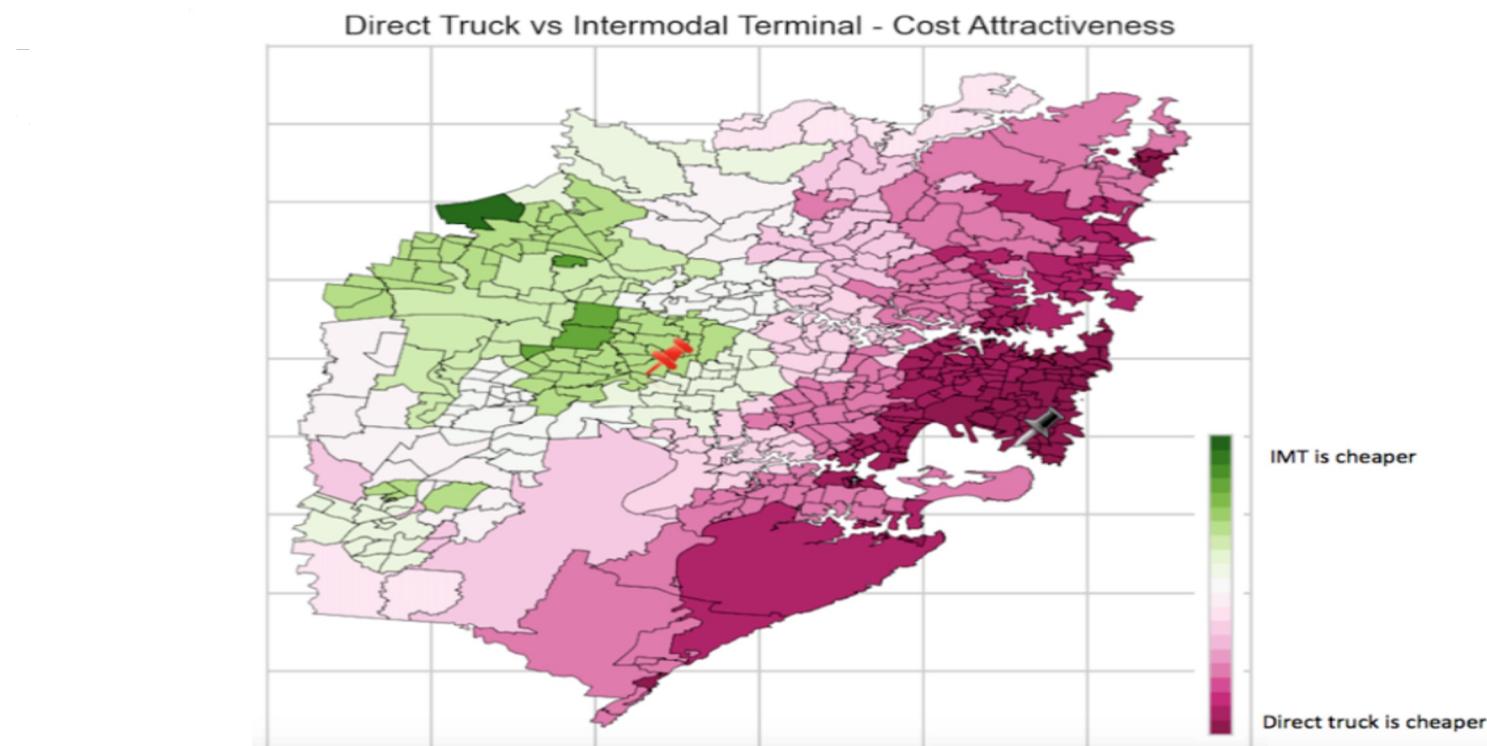


Figure 2. Map of regions coloured by cost difference between using IMT (red pin) vs. direct truck to port Botany (black pin).

Using machine learning in supply chains

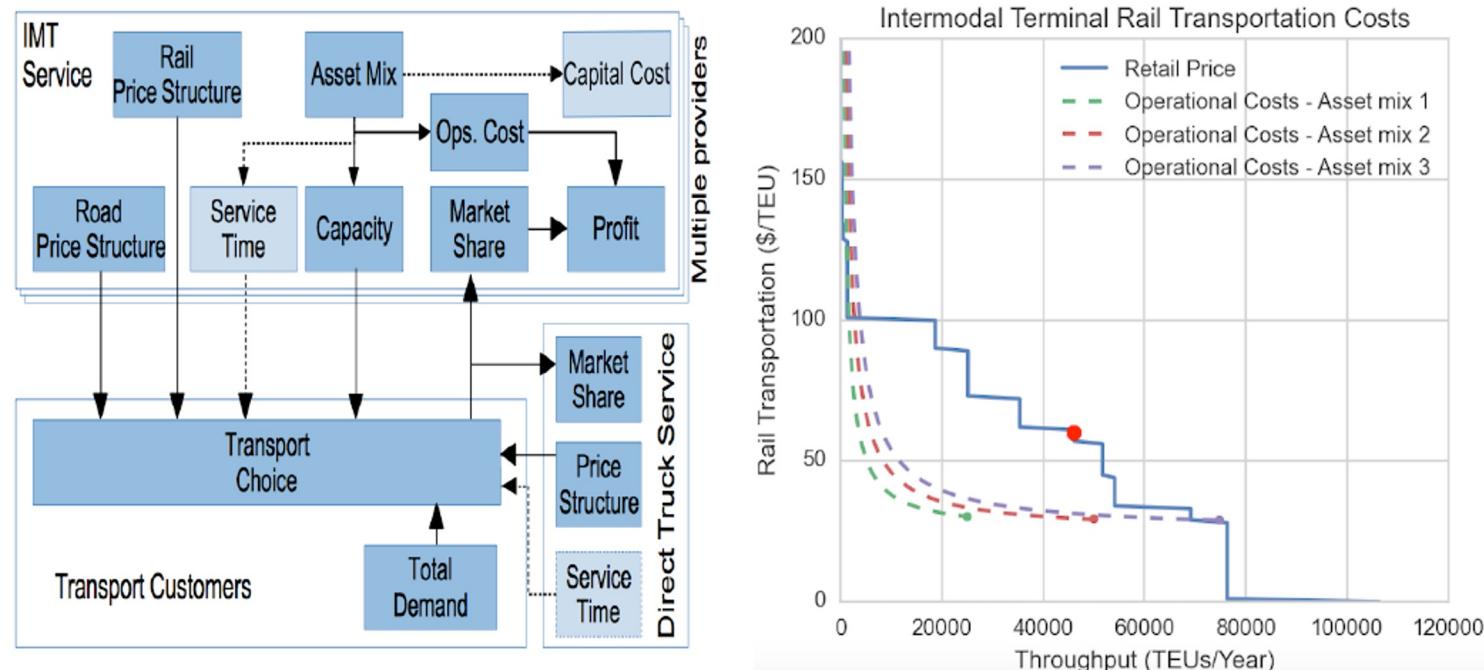


Figure 1. (Left panel) Schematic of the model. (Right panel) Plot of retail rail price (blue) and operational costs incurred by IMTs (dotted lines) vs. container throughput via the IMT. The red dot marks the optimal retail rail price for Asset mix 2 that maximizes the profit of the IMT.

Explainable AI

The case of lithium

- Lithium equation
- Look into data
- Sources of data
- Explained simpler models first
- Saga of data access
- Validation