

# Welcome to the LLMs for Bio hackathon, aboard the Gene Chaser

The background image shows a large cruise ship, the "Gene Chaser", docked at a port. The ship is white with blue accents and has several decks visible. It is positioned in front of a city skyline at dusk or night, with tall buildings and lights reflected in the water. The sky is dark with some clouds.

Innovating at the intersection of **Generative Artificial Intelligence** and **Life Sciences + Health**



SGInnovate



4Catalyzer



Lambda



## A word from Dr. Jonathan Rothberg

Inventor of high-speed Next-Gen DNA sequencing  
Adjunct Professor of Genetics at Yale University

# Hackathon Team Intros



Dr. Vicky Demas  
CEO at Identifeye HEALTH

# AI intersections at the 4C ecosystem!

Vicky Demas  
CEO identifeye HEALTH  
July 28, 2023

# AI at an inflection point

1997  
Deep Blue defeats Garry Kasparov in chess match

WIRED STAFF SCIENCE JUN 26, 2012 11:15 AM

2011: IBM's Watson wins Jeopardy



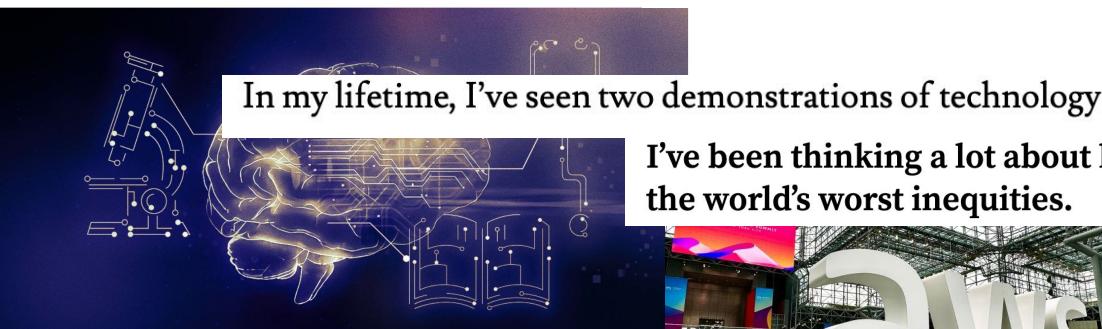
A NEW ERA

## The Age of AI has begun

Artificial intelligence is as revolutionary as mobile phones and the Internet.

By Bill Gates | March 21, 2023

the initial release of the iPhone which transformed what a phone actually was, and gave rise to the modern 'smartphone'.



Attendees learn how generative AI is transforming organizations, across all kinds of industries and applications.

"Generative AI has captured our imaginations," Sivasubramanian said.

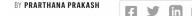
"This technology has reached its tipping point."

## FORTUNE

TECH - GOOGLE

Alphabet CEO Sundar Pichai says that A.I. could be 'more profound' than both fire and electricity—but he's been saying the same thing for years

BY PRARTHANA PRAKASH April 17, 2023 at 10:50 AM PDT



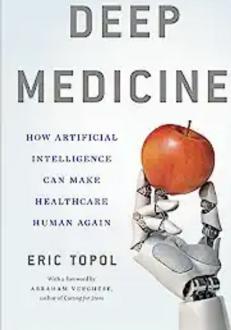
In my lifetime, I've seen two demonstrations of technology that struck me as revolutionary.  
I've been thinking a lot about how AI can reduce some of the world's worst inequities.

## nature

NEWS FEATURE | 25 July 2023

## ChatGPT broke the Turing test – the race is on for new ways to assess AI

Large language models mimic human chatter, but scientists disagree on their ability to reason.



# 4Catalyzer



4Catalyzer is a technology incubator founded by Dr. Jonathan Rothberg with the **mission to save lives and maximize societal impact.**

We work at the intersection of engineering, machine learning and natural sciences to support innovation in healthcare and life sciences.

4Catalyzer companies have been named Technology Pioneers three times by the World Economic Forum and top innovations of the year twice by peer-reviewed academic journal Science Magazine. 4Catalyzer has raised over \$2 billion in funding and taken three companies public.

# 4Catalyzer



HEALTHCARE



**HYPERFINE**



**Detect**



Liminal Sciences

LIFE SCIENCE  
TOOLS



CLIMATE TECH

**PROTEIN  
EVOLUTION**

# Why are we here?



Creating tools to:

- Help patients (or help doctors help patients)
- Establish knowledge/ground truth
- Solve societal issues

## Healthcare is often inaccessible to many patients



### **Confined to specialists' offices**

Imaging and testing is primarily confined to urban settings, large hospitals and specialists' offices, making access to care inconvenient and difficult for patients.



### **Large and complex devices, complex workflows**

Existing devices are immobile or complex and don't allow providers to meet patients where they are



### **Prohibitively expensive**

The price point of existing devices puts them out of reach for most non-specialized facilities and use cases

## Drug Development is costly and Drugs can be inaccessible to patients

**Lengthy, complex, and costly** – high degree of uncertainty that a drug will actually succeed.

**Target Identification is challenging** – unknown pathophysiology for many disorders

**Heterogeneity of the patient population**  
-requires processing of large amounts of data  
(clinical phenotyping and subtyping)

**Lack of tools** for improved target identification and validation

## Life science tools are often inaccessible to many researchers



### Confined to specialized labs

testing is primarily confined to specialized labs and



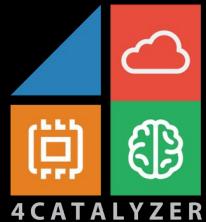
### Large and complex devices

Existing devices are complex, have large physical overhead



### Prohibitively expensive

The price point of life sciences tools such as sequencing platforms and mass spectrometers (including reagents and test kits and data analytics) puts them out of reach for most non-specialized facilities and use cases



# Butterfly™

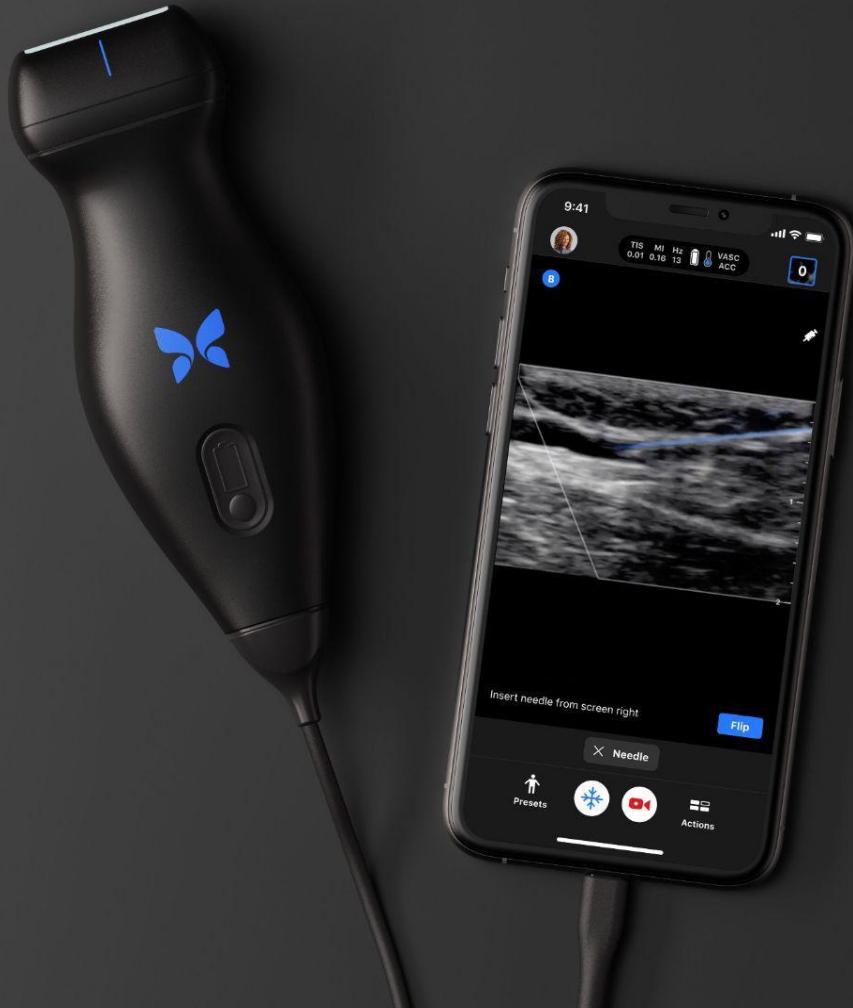
Butterfly is commercializing the world's first ultrasound on a semiconductor chip. "BFLY" went public in February of 2021 on the NYSE raising half a billion dollars to continue to increase hospital connectivity as well as to further develop AI applications, ultimately enabling home use.

**NYSE: BFLY**

# Democratizing access to Ultrasound Imaging

- World's first full body ultrasound scanner on a semiconductor chip, fused with AI and cloud technology
- The quality of a \$60k device for a fraction of the price
- Two thirds of the world does not have access to medical imaging, and two thirds of diagnostic issues can be resolved with simple imaging.
- Saving lives on all 7 continents, partnership with the Gates Foundation for large scale deployments in Limited Resource Settings

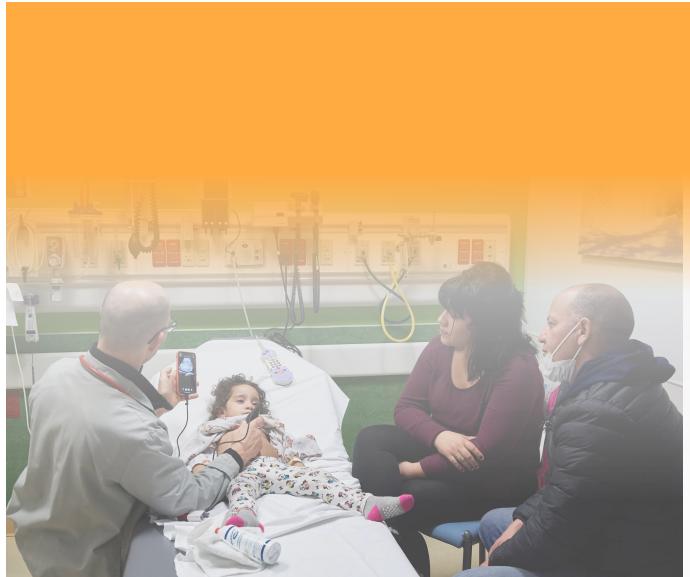






2/3

of diagnostic dilemmas  
can be solved with  
medical imaging



2/3

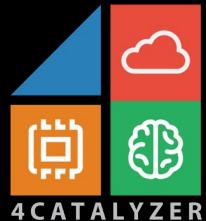
of the world has  
no access to  
medical imaging



100+ Global Health Partners.  
50+ Low Resource areas.

Hope in the Palm of a Hand  
Largest Global Health Deployment of  
Ultrasounds in history





# HYPERFINE

Hyperfine has developed the world's first portable MRI device that can be used anytime, anywhere. "HYPR" went public in 2021 on NASDAQ raising quarter billion dollars to accelerate commercialization as well as to further leverage AI to improve images.

**NASDAQ: HYPR**

# Democratizing access to MRI

- MRI is one of the safest forms of medical imaging - but is unavailable in 90% of the world - Hyperfine is changing that
- World's first Portable MRI
  - Low Cost, easy to transport to the patient's bedside, plugs into a wall outlet
- Powered by AI to give clear images and insights
- In use at over 100 sites worldwide





NOTICE  
CAUTION - Dangerous  
Limited to professional use by  
investigation firms

**HYPERFINE**

EMERGENCY  
MEDICAL SERVICE

**LIFELINE**  
AMBULANCE SERVICE

718-824-4500

**Meet Swoop™**

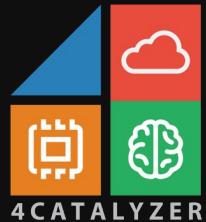
Portable MR Imaging.

Here. Now.

**HYPERFINE**

[www.hyperfine.io](http://www.hyperfine.io)  
@hyperfineR (866) SWOOP-MR





# Liminal Sciences

Liminal Sciences is developing a non- invasive brain monitor to sense, understand, and ultimately heal the brain.

# Liminal is building the world's first non-invasive brain monitor

- The Brain is as important to monitor as the heart – but today there is no way to do this without drilling a hole in the patients skull
- Liminal is building the first wearable brain monitor for acute and chronic conditions including epilepsy, stroke, and traumatic brain injury.
- Enabling a brain monitor as ubiquitous as the heart monitor
- Partnered with AI for quicker and more valuable health insights.





Liminal

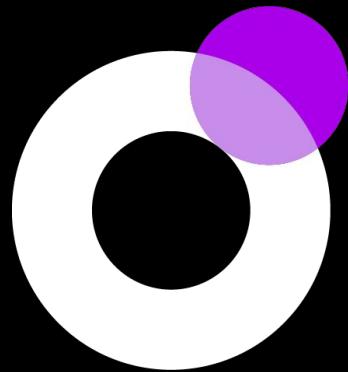
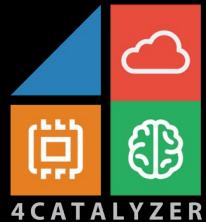
LPC Rev. 1

SN:



Dr. David Rosenthal MD  
Internal Medicine





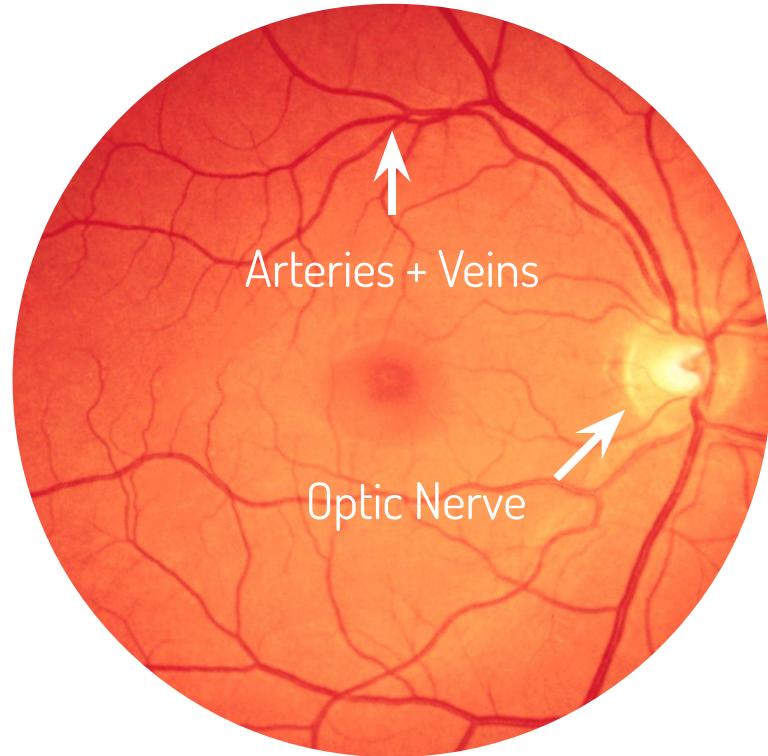
# identifeye

[HEALTH]

identifeye HEALTH (fka Tesseract Health) raised \$80M through Series B to develop and launch its first product. We are building an intuitive, consumer friendly, medical device to capture health information from the eye.

HARNESSING THE POWER OF THE EYE

A window into the body  
and health



THE RETINA

# An abundance of health information



Ocular Diseases



Neurological Diseases



Cardiovascular Diseases



Infectious Diseases



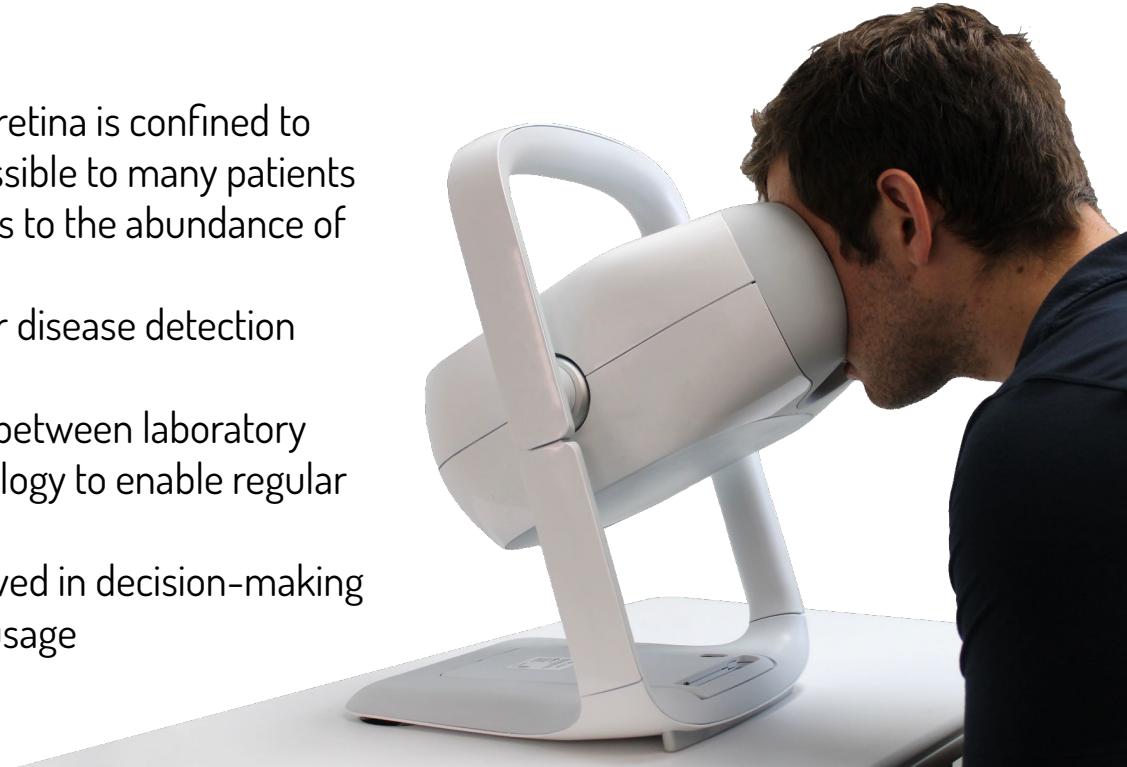
Inflammatory Diseases



Blood tests + Vital Signs

# Create a New Branch in Diagnostics - Making Retinal Imaging as simple as an eye selfie

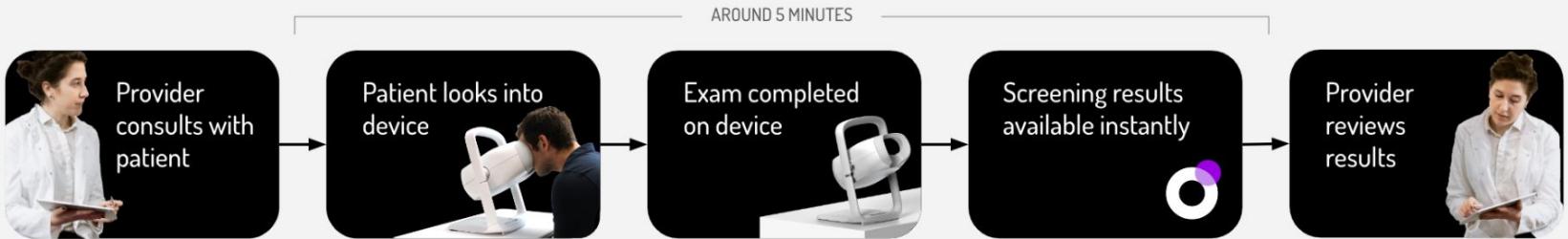
- Access to health information in the retina is confined to specialists' offices and often inaccessible to many patients
- Our mission is to democratize access to the abundance of health information in the eye
- Prevent vision loss by making ocular disease detection easier and more accessible
- Create a new branch in diagnostics between laboratory medicine (e.g. bloodwork) and radiology to enable regular non-invasive health monitoring
- Empower patients to be more involved in decision-making by reducing barriers to access and usage



# Some Examples of Leveraging AI to solve problems at identifeye

- Remove the need of trained/specialized operators by leveraging AI to automate adequate quality images
- Improve/increase the training datasets, e.g.
  - simulate photos from different camera systems
  - generate images with representative features spanning disease burden
  - Clean up data /e.g. case report forms
- Develop disease classifiers - starting with Diabetic Retinopathy
  - Human crafted “relevant” features
  - Black Box approach
- Create custom training materials
- Workflow optimization
- ++

# Automating Device Operation



## 1. Instruction

- Expectation setting
- Patient positioning + adjustments
- Guidance on where to look

## 2. Image Capture

- Auto-alignment
- Auto-focus
- Auto-exposure
- Auto-capture

## 3. Quality Assessment

- Automated image quality assessment
- Guidance for image recapture if necessary

## 4. DR Screening

- Automated screening for diabetes-related retinal disease features

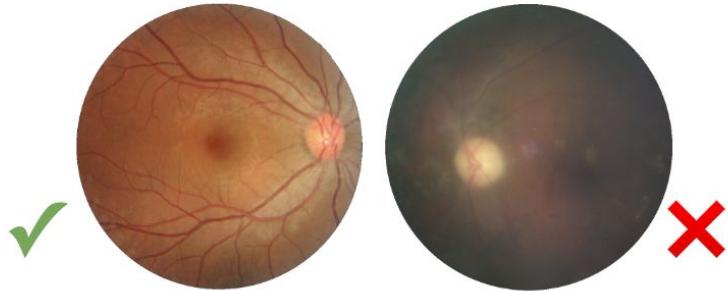
## 5. Result Transmission

- Immediate + secure storage of results and images
- Instant transmission of results to EMR

# Image Quality Assessment and Disease Diagnosis

## Image Quality Assessment (IQA)

Ensure high-quality data

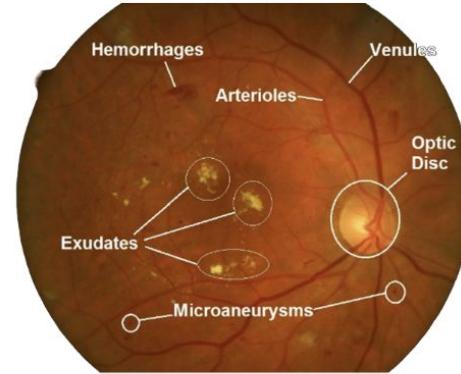


Have we captured a quality image for clinical interpretation?

- Focus / sharpness
- Illumination uniformity, contrast
- Color balance
- Debris, artifacts, obstruction
- Haze / glare
- Is the right region captured?

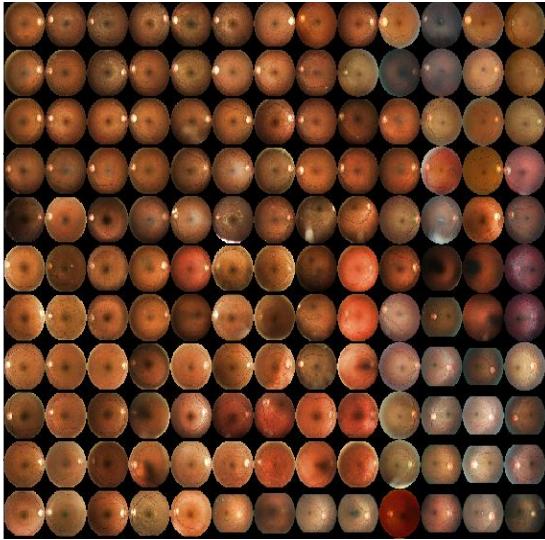
## Diabetic Retinopathy Screening (DR)

Make it actionable



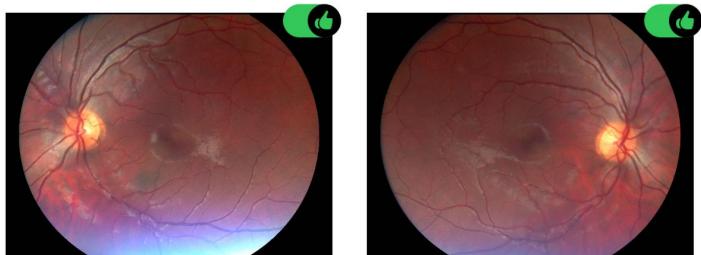
- Instantly detect retinal disease features associated with diabetes
- Refer concerning cases to specialists
- Avoid further retinal damage
- Help prevent blindness

# Algorithm Development

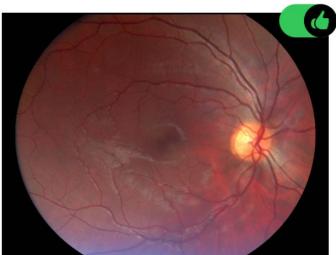


- Training on tens of thousands of images from existing databases
- Assortment of cameras, FOVs, centration, quality
- 5 class labels + Ungradable, No DR, Mild, Moderate, Severe, Proliferative
- Predict referable (more-than-mild) vs. non-referable
- Iterative tuning with thousands of images from our development and clinical units
- Eventually validation

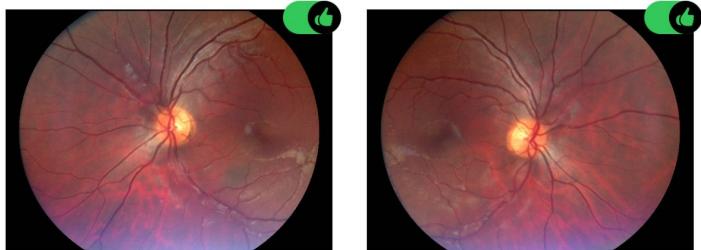
# Automating Device Operation -- Image Capture



Macula Centered (Left)



Macula Centered (Right)



Disc Centered (Left)



Disc Centered (Right)

## Negative

Return for screening  
in 12 months



## Positive or Ungradable

Refer to a specialist



# AI in Healthcare

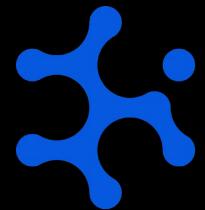
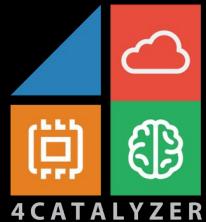
REVOLUTIONIZING THE WHERE, WHEN, WHAT AND HOW PEOPLE ACCESS HEALTHCARE



# AI in Healthcare

## REVOLUTIONIZING THE WHERE, WHEN, WHAT AND HOW PEOPLE ACCESS HEALTHCARE

- Remove barriers and accessibility to skilled healthcare providers
- Data Driven Clinical Decision Support for Disease Diagnosis and Patient Monitoring
- Image analysis for early detection and diagnosis of disease (e.g. image segmentation, anomaly detection)
- Hospital Management and Optimization
- Digital Consultation, virtual Assistants and chatbots - e.g. appointment scheduling, medication and appointment reminders, patient triaging
- Treatment Design - personalized treatments based on individual's genetic and clinical data, environmental and behavioral
- Drug Discovery and Development - analyze chemical and biological data to identify drug candidates, optimize design, predict efficacy and safety

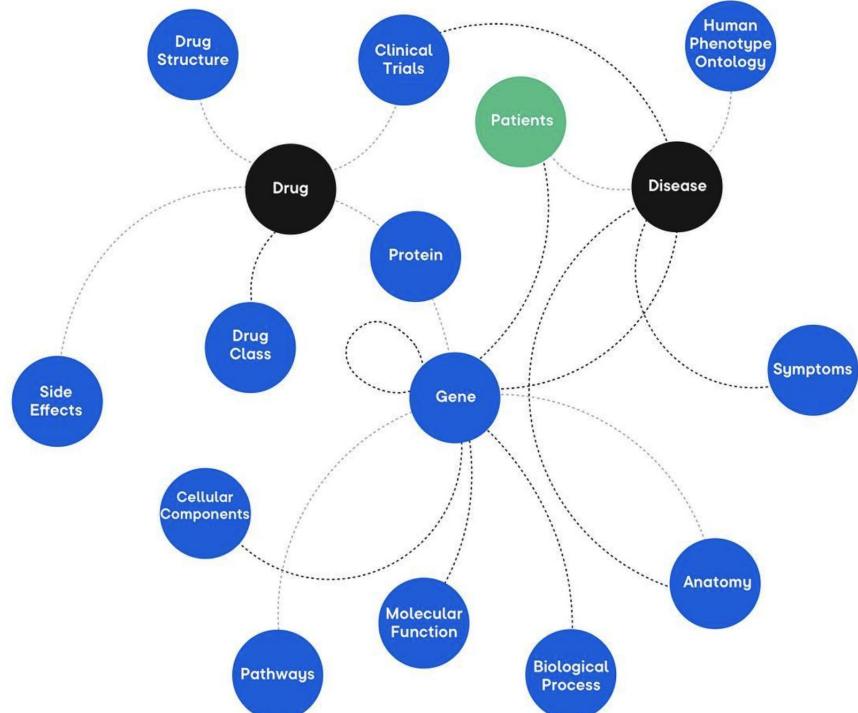


# AI Therapeutics

AI therapeutics is identifying and developing drugs  
& treatments for rare orphan conditions.

# Identifying and developing drugs for rare conditions

- Uses AI Algorithms to identify promising assets to add to their pipeline
- AI analyzes what these safe drugs can be used for
- Helping find treat for orphan/rare diseases
- Current clinical stage drug candidates target several orphan disorders including amyotrophic lateral sclerosis, pulmonary arterial hypertension, bronchiolitis obliterans, pulmonary sarcoidosis, Ewing sarcoma, rhabdoid tumor, and SWI/SNF mutated or dysregulated cancers.



# AI in Life Sciences

CREATING BETTER TOOLS TO STUDY AND UNDERSTAND THE WORLD AROUND US

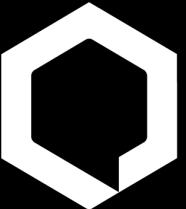
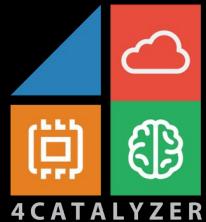


Antonie van Leeuwenhoek (1632–1723) was one of the first people to observe microorganisms, using a microscope of his own design, establishing that there were forms of life that were not visible to the naked eye.

# AI in Life Sciences

**CREATING BETTER TOOLS TO STUDY AND UNDERSTAND THE WORLD AROUND US**

- Data Analysis and Interpretation – analyze large volumes of data to identify patterns and extract insights (e.g. next generation sequencing, microscopy and imaging)
- Technology / platform development
- Laboratory Automation – optimize experimental workflows, experiment planning
- Antibody Design and Optimization
- Protein Engineering – desired properties



# QuantumSi

Quantum-Si has built the world's first single molecule protein sequencer on a semiconductor chip. QSi went public in 2021 raising over \$500 million to revolutionize the future of diagnostics.

**NASDAQ: QSI**

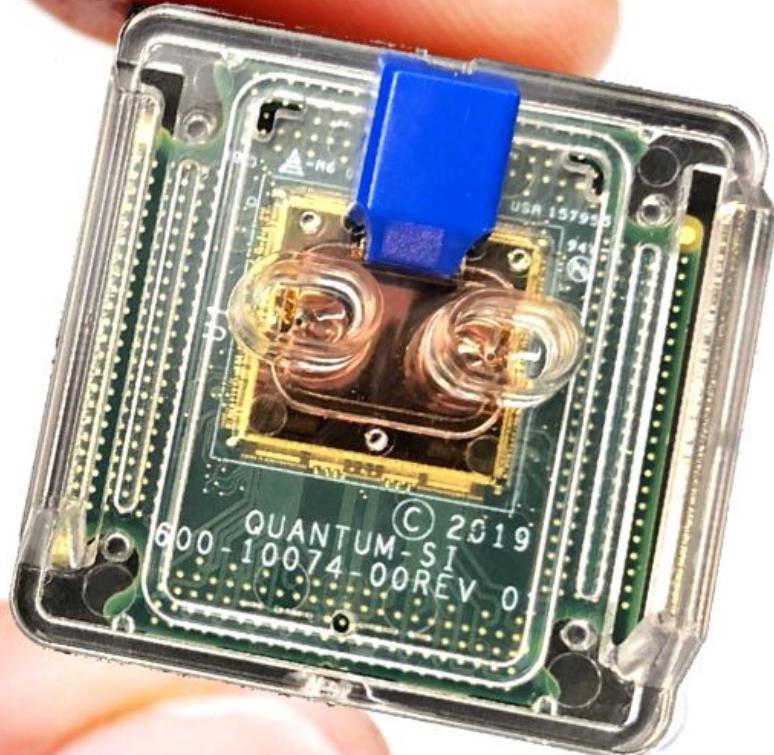
# Access to rapid PCR quality testing

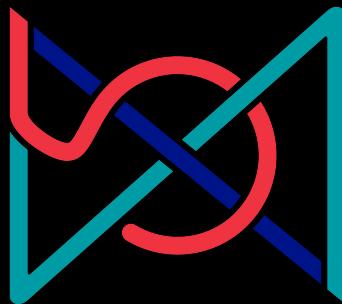
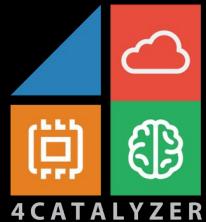
- First ever single molecule protein sequencing device
- DNA tells you what may happen, proteins tell you what is happening/about to happen
  - Could give us the ability to predict a heart attack before it happens with high accuracy
- Deeper proteomic insights will advance science and human health



# Next-Gen Protein Sequencing







# 454 Bio

Founded in 2022, 454 Bio is building the first Next Generation DNA sequencer (NGS) fit for at-home use

# Decentralized On-site Sequencing

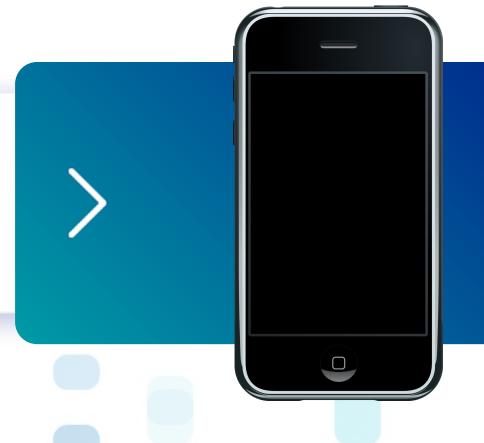
- Today, personal genomics is limited by expensive and cumbersome equipment - turnaround time is slow, and sequencing requires expertise and resources
- 454 Bio's mission is to bring genomics out of the lab and enable universal access to affordable and fast DNA sequencing.
- Low Cost device, easy to use kits, seamless data transfer for results without expertise
- 454 Bio is supporting a wide range of use cases in public health, research, and consumer genomics.



# 454 Bio: We need a personal genomics revolution



Personal computing revolution



Personalized healthcare revolution



# The 454 sequencer is the most cost effective and easy to use NGS machine ever made



## Device

Low cost device



## Consumables

Low cost kits

## Key Features:



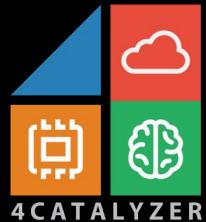
Low cost device (\$99)\*



No specialized facilities, personnel or accessory equipment needed to operate



Rapid results



# Detect

In response to the global pandemic, Detect built and brought-to-market a PCR-quality rapid molecular home Covid-19 test authorized for EUA by the FDA. It is now focusing on a platform for POC and home use.

# Access to rapid PCR quality testing

- Detect's proprietary technology serves as the platform for its future home tests, including the next-generation Detect Covid+Flu Test, as well as rapid molecular home tests for respiratory health, Strep and STIs.
- Super easy to use, low cost, and fast
- Enables stakeholders to test on site, without sending to a lab



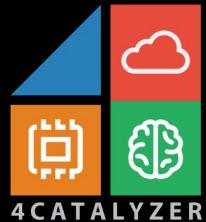






Detect

Detect



# PROTEIN EVOLUTION

Protein Evolution is leveraging recent breakthroughs in natural science and artificial intelligence to design enzymes to break down end-of-life textile and plastic waste into the building blocks that make up new textile and plastic products. Protein Evolution aims to help the chemicals industry transition to a lower-carbon, circular economy.

380 million tonnes of plastic  
produced per year<sup>1</sup>

PROTEIN  
EVOLUTION



92% enters landfills, oceans or is  
incinerated<sup>2</sup>

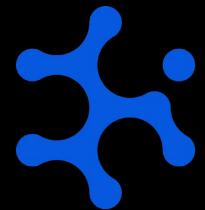
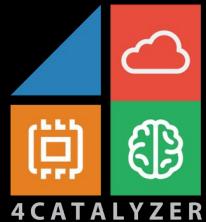
1. Source: Plastics Europe, 2020

2. Source: Plastics Europe, 2020

# Infinitely recyclable High quality plastic

- Identify the waste - water bottle, car tire, or piece of clothing.
- Engineer enzymes - Uniting natural science and artificial intelligence, develop enzymes that break down the waste source so we can recycle it in an economical, sustainable way.
- Break down the plastic into its “building blocks” in a low-emission, eco-friendly process.
- Reproduce materials by using these “building blocks” to create good-as-new plastic bottles, textiles, and other infinitely renewable plastic products.



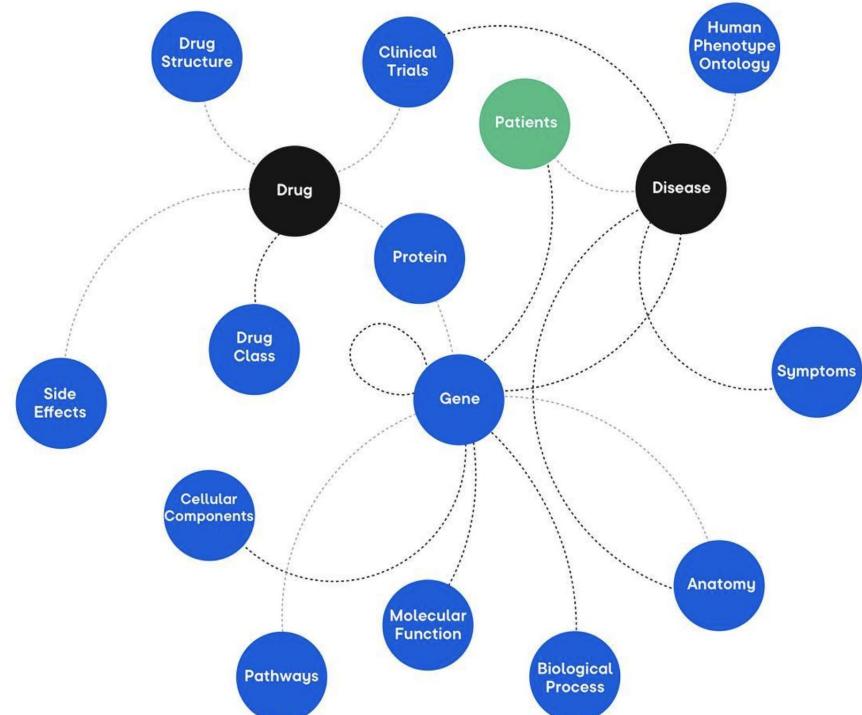


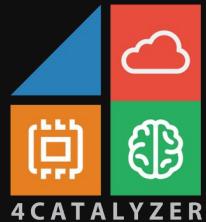
# AI Therapeutics

AI therapeutics is identifying and developing drugs  
& treatments for rare orphan conditions.

# Identifying and developing drugs for rare conditions

- Uses AI Algorithms to identify promising assets to add to their pipeline
- AI analyzes what these safe drugs can be used for
- Helping find treat for orphan/rare diseases
- Current clinical stage drug candidates target several orphan disorders including amyotrophic lateral sclerosis, pulmonary arterial hypertension, bronchiolitis obliterans, pulmonary sarcoidosis, Ewing sarcoma, rhabdoid tumor, and SWI/SNF mutated or dysregulated cancers.





# Liminal Sciences

Liminal Sciences is developing a non- invasive brain monitor to sense, understand, and ultimately heal the brain.

# Liminal is building the world's first non-invasive brain monitor

- The Brain is as important to monitor as the heart – but today there is no way to do this without drilling a hole in the patients skull
- Liminal is building the first wearable brain monitor for acute and chronic conditions including epilepsy, stroke, and traumatic brain injury.
- Enabling a brain monitor as ubiquitous as the heart monitor
- Partnered with AI for quicker and more valuable health insights.



# AI and Big Data - Magnify our potential to save lives and maximize societal impact

“The best way to predict the future is  
to make it”  
Dr Jonathan Rothberg



**Dr. Stefan Harrer**  
Chief Innovation Officer at Digital Health CRC

2023 Hackathon: Large Language Models for Bio

# Large Language Models in Health and Medicine – trends, ideas, opportunities

**Dr. Stefan Harrer**

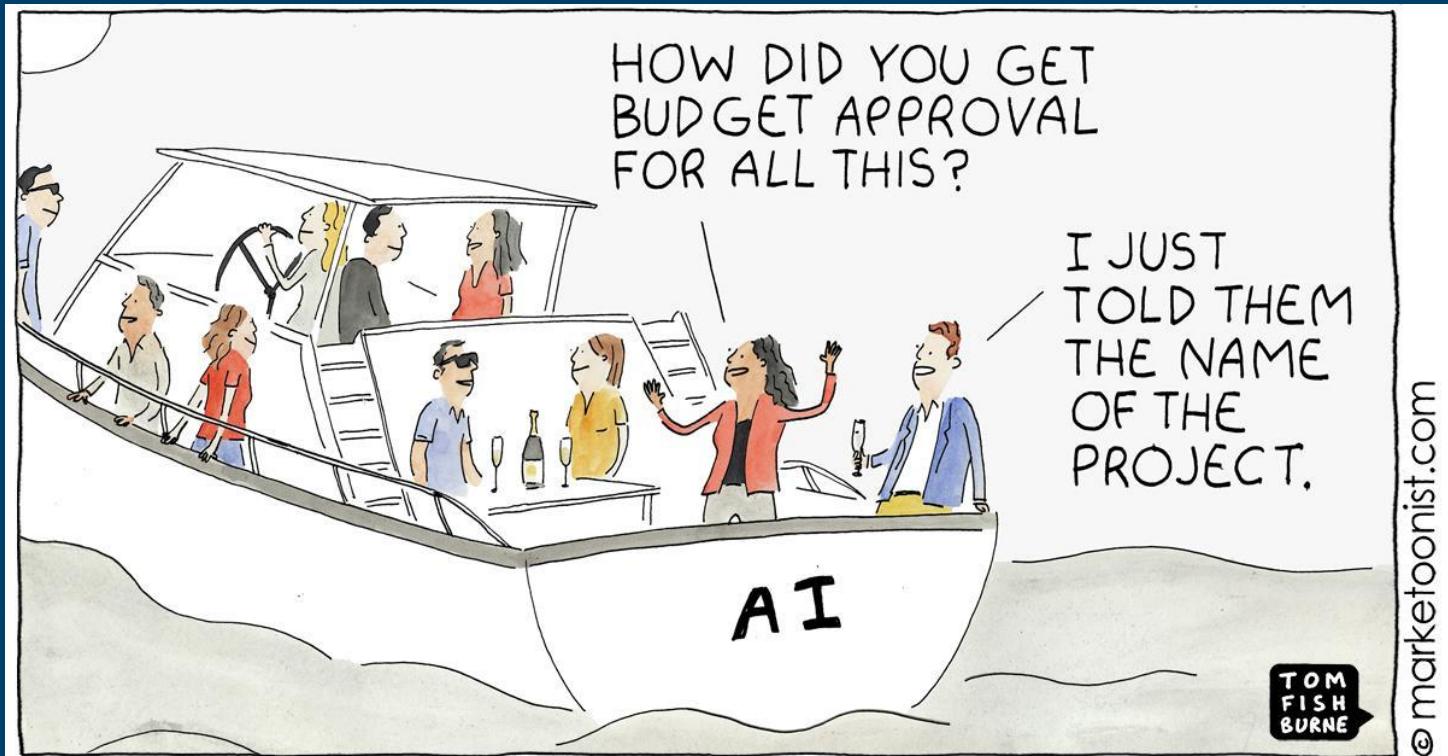
Chief Innovation Officer, Digital Health Cooperative Research Centre

Adjunct Professor, University of Technology Sydney

Member Forbes Technology Council, IEEE Senior Member, Member New York Academy of Sciences

July 28, 2023

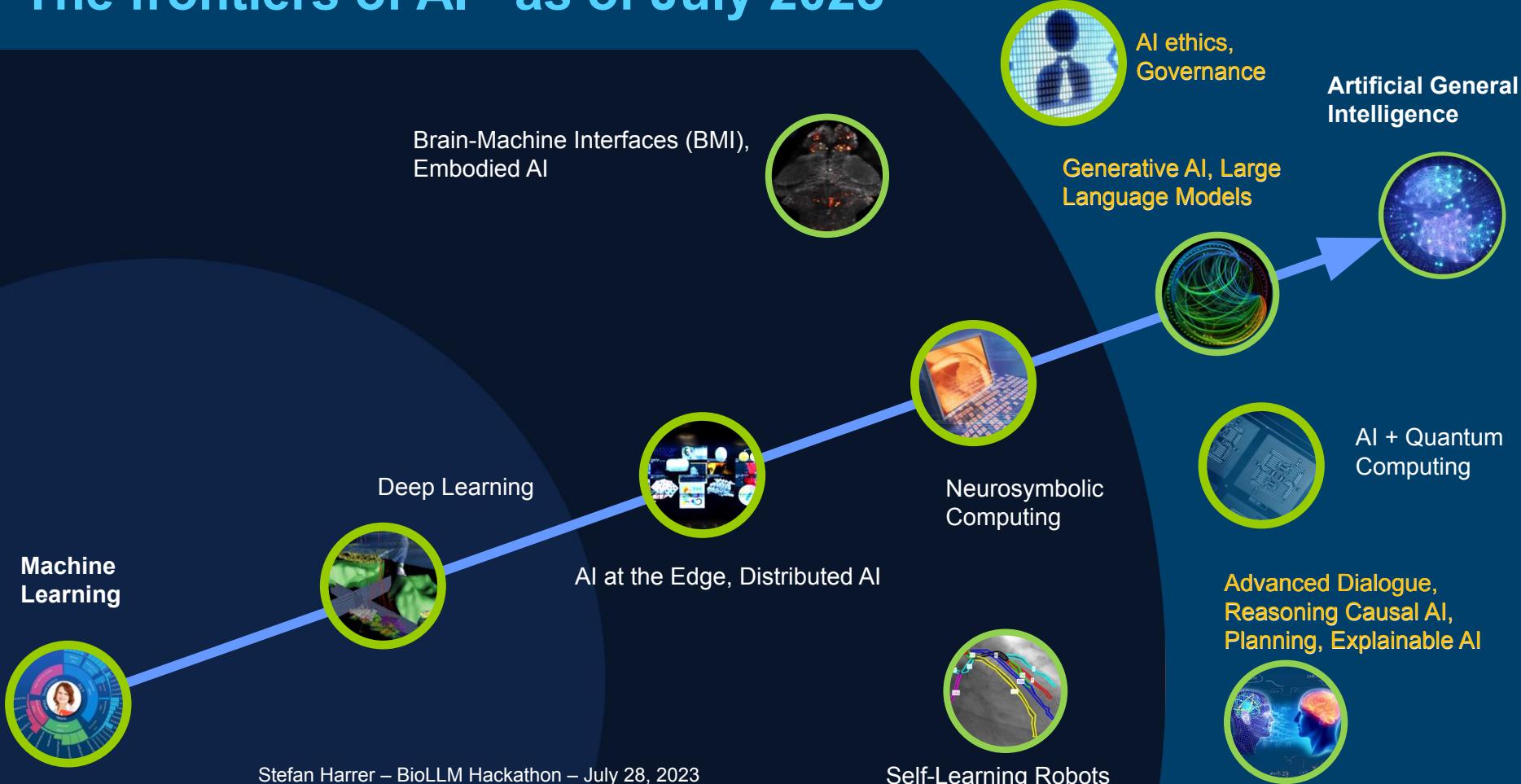
# Not everything with AI on it has AI in it



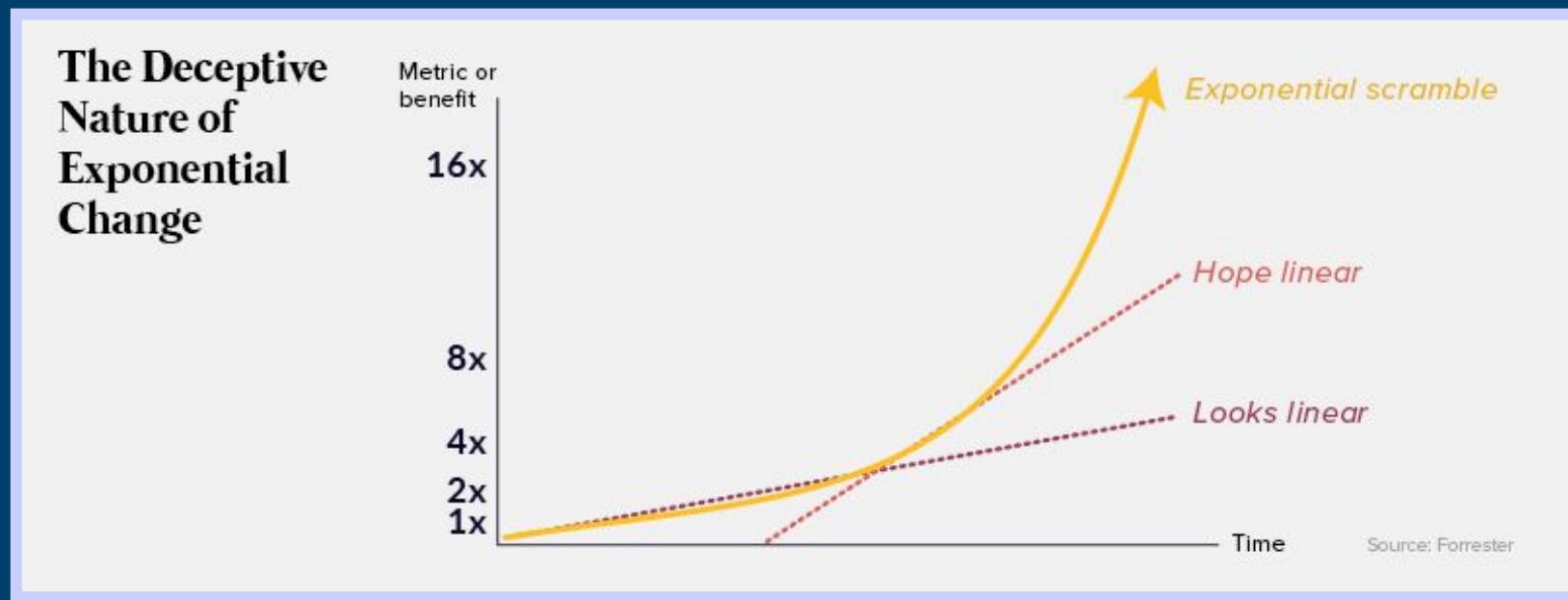
# Intelligence is a complex concept



# The frontiers of AI - as of July 2023

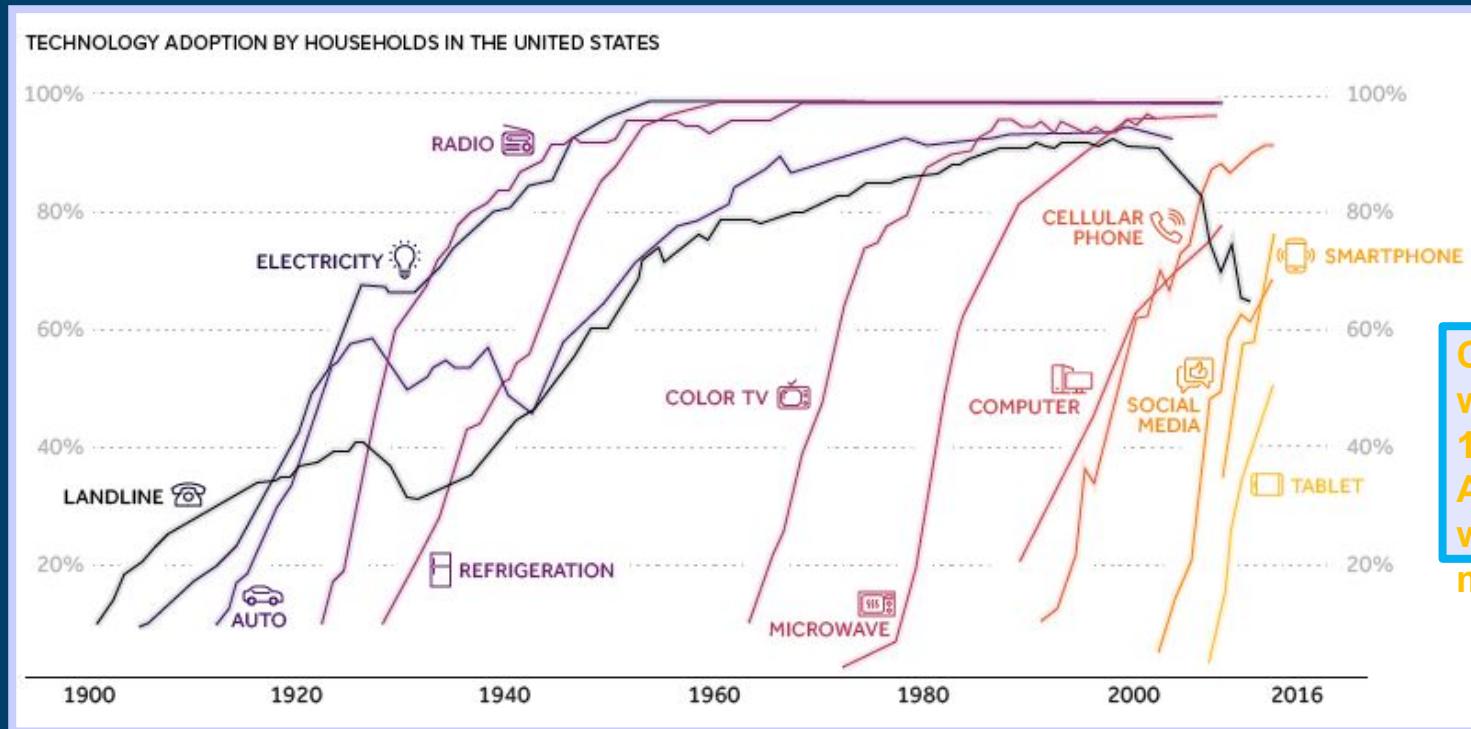


# The pace of technological progress – a human misconception



<http://www.visualcapitalist.com/the-8-major-forces-shaping-the-future-of-the-global-economy/>

# The future comes earlier these days than it used to...



ChatGPT went to 100M Daily Active Users within two months

<http://www.visualcapitalist.com/the-8-major-forces-shaping-the-future-of-the-global-economy/>

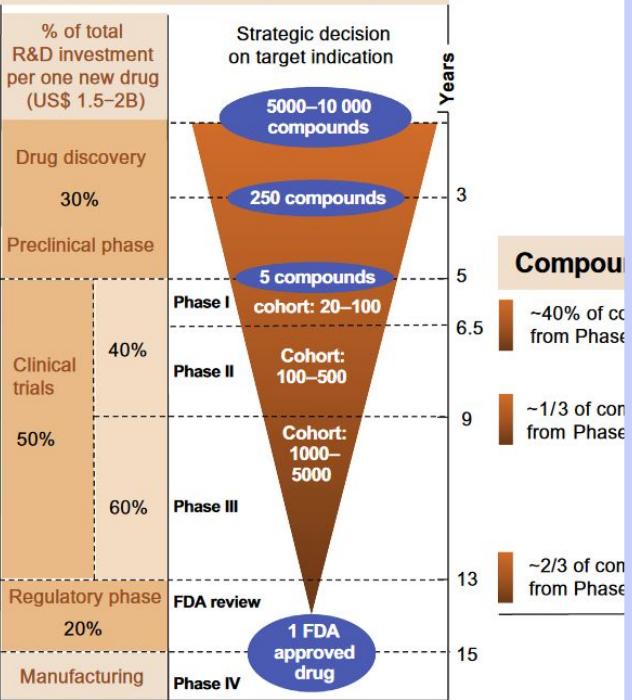
# How can we use Large Language Models in health and medicine?

- assist clinicians with the generation of medical reports, information retrieval, documentation
- simplify medical jargon in clinician–patient communication
- increase the efficiency of clinical trial design
- help to overcome interoperability and standardisation hurdles in EHR mining
- make drug discovery, design and development processes more efficient

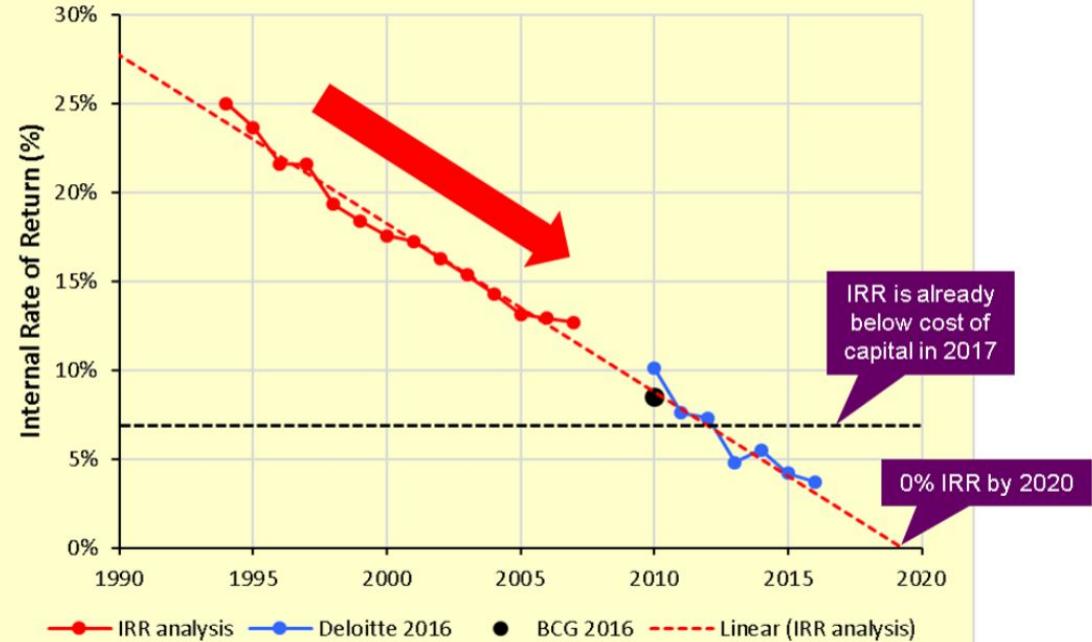
The screenshot shows the MedPage Today homepage with a navigation bar at the top. Below the navigation, there's an advertisement for a medical word game called "WORDOSIS". The main headline reads "Google AI Performs at 'Expert' Level on U.S. Medical Licensing Exam". A subtext below the headline states "Latest version of medically tuned AI model achieved 85% accuracy, beating previous record". The author of the article is Michael DePeau-Wilson, Enterprise & Investigative Writer, MedPage Today, dated March 14, 2023.

# LLM use case: drug discovery, development – Eroom's Law

## Pharma drug development cycle

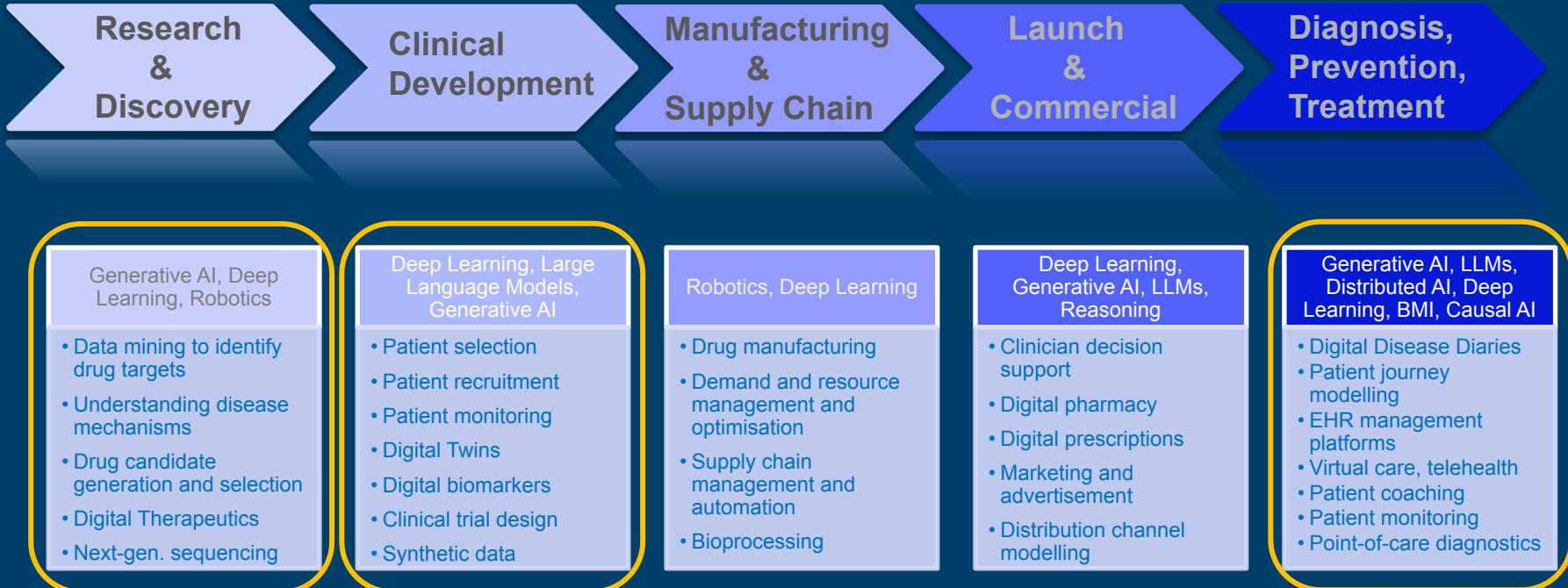


## Return on Investment in Pharma R&D



S. Harrer et al., 'Artificial intelligence for clinical trial design', *Trends in Pharmacological Sciences*, CellPress, 2019.

# (Generative) AI transforms pharma and biotech



# AlphaFold: an AI breakthrough eyeing the Nobel Prize

- In 2020 Google DeepMind Introduces AlphaFold, a deep learning system that can accurately predict 3D models of protein structures. AlphaFold has solved the 50-year-old “protein folding problem” and marks a watershed moment for drug discovery.
- By July 2022 DeepMind had expanded the openly available AlphaFold Protein Structure Database from nearly 1 million to over 200 million structures, including predictions for most proteins in UniProt, thus including the structure of nearly every protein known to mankind.

Article | [Open Access](#) | Published: 15 July 2021

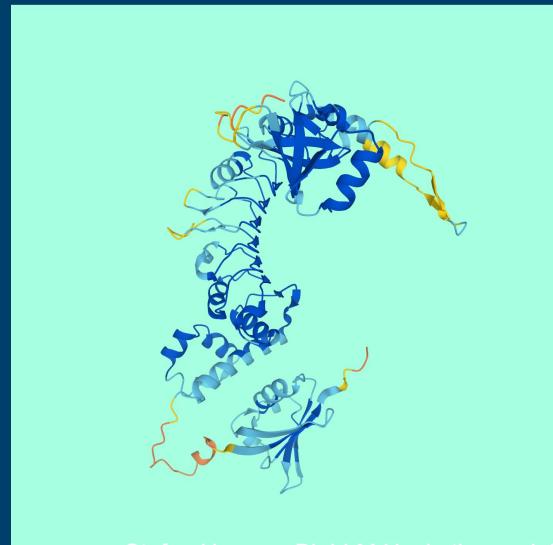
## Highly accurate protein structure prediction with AlphaFold

John Jumper , Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, Alex Bridgland, Clemens Meyer, Simon A. Kohl, Andrew J. Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislav Nikolov, Rishabh Jain, Jonas Adler, Trevor Back, Stig Petersen, David Reiman, Ellen Clancy, Michal Zielinski, ... Demis Hassabis 

+ Show authors

*Nature* **596**, 583–589 (2021) | [Cite this article](#)

**1.13m** Accesses | **7323** Citations | **3435** Altmetric | [Metrics](#)



# The perfect storm: add LLMs, generative AI

- In February 2022, InSilico Medicine reported the start of Phase I clinical trials for the first-ever AI-discovered molecule based on an AI-discovered novel target—all done at a fraction of the time and cost of traditional preclinical programs.
- In January 2023, AbSci became the first entity “to create and validate *de novo* antibodies *in silico*” using generative AI.
- In February 2023, the FDA granted its first Orphan Drug Designation to a drug discovered and designed using AI;
- In June 2023, InSilico Medicine commences global Phase II trial for the drug.

2022.02.24

## **From Start to Phase 1 in 30 Months: AI-discovered and AI-designed Anti-fibrotic Drug Enters Phase I Clinical Trial**



# LLMs can improve clinical trial efficiency: endpoint detection, adherence control, retention risk prediction



# LLMs can improve clinical trial efficiency: digital twins and synthetic control arms

Clinical Trials Arena |  

News Analysis Sectors Themes Insights Companies Events Reports  Premium Insights

News | June 30, 2023

## QurAlis and Unlearn.AI collaborate to optimise generative AI-powered ALS clinical trials

QurAlis will leverage Unlearn.AI's machine learning algorithms to create digital twins of patients enrolled in clinical trials for amyotrophic lateral sclerosis (ALS).

By Elizabeth Cooke

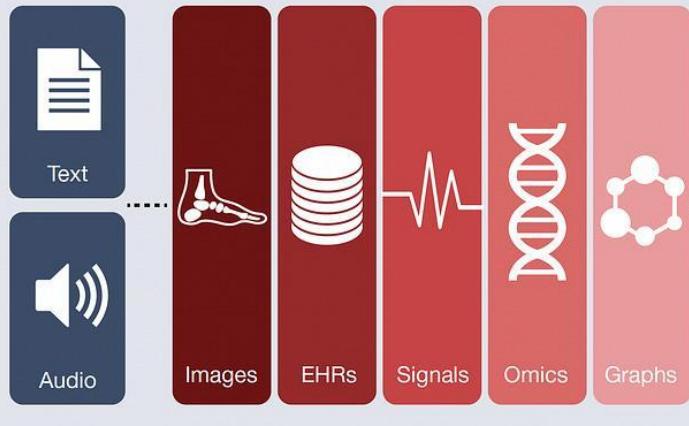


# LLM use case: generalist medical AI

Stefan Harrer – BioLLM Hackathon – July 28, 2023

a

## Multimodal self-supervised training

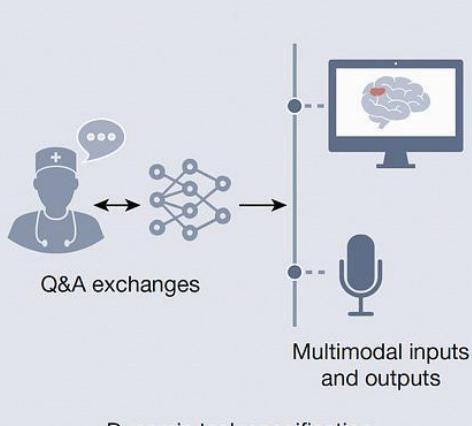


## Medical domain knowledge



Reasoning with multiple knowledge sources

## Flexible interactions



Reasoning with multiple knowledge sources

b

## Applications



Chatbots for patients



Interactive note-taking



Augmented procedures

...



Grounded radiology reports



Text-to-protein generation

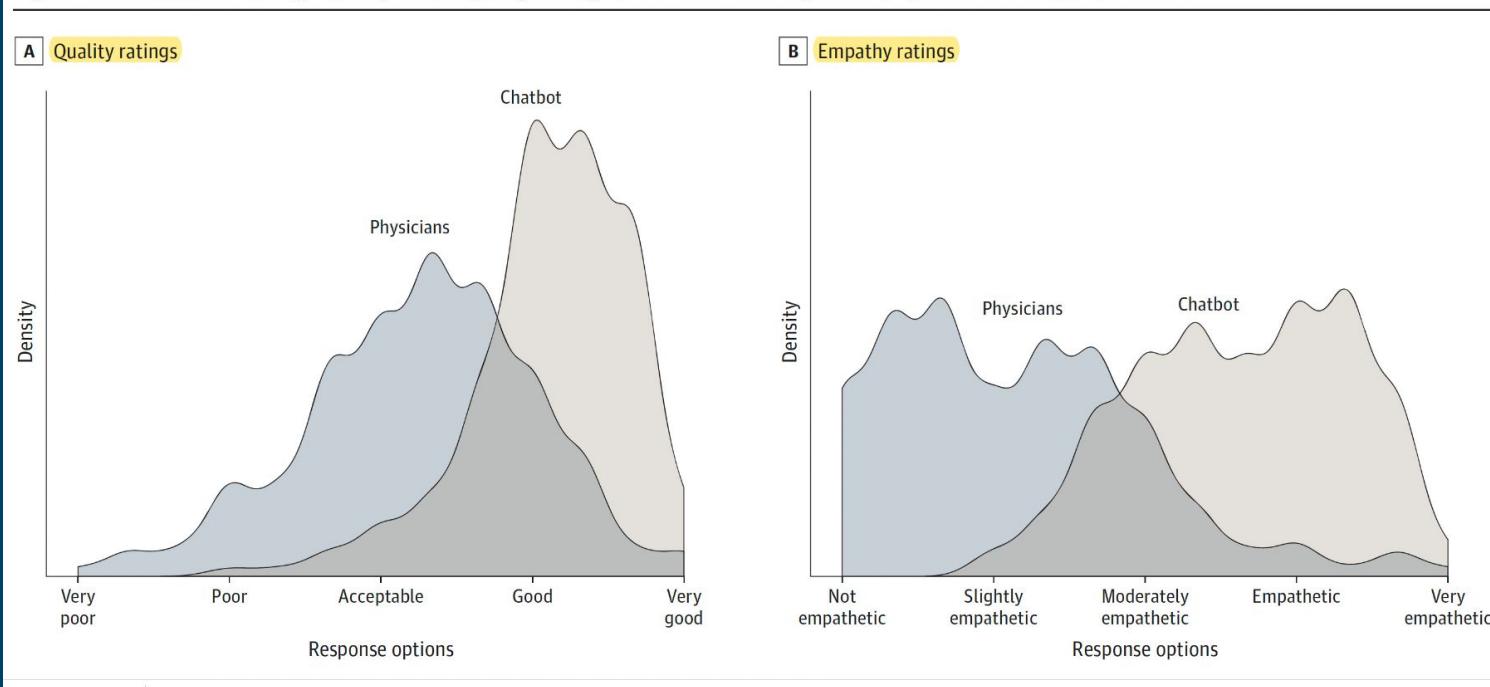


Bedside decision support

**Regulations:** Application approval; validation; audits; community-based challenges; analyses of biases, fairness and diversity

# LLMs for clinician-patient interactions, education, coaching

Figure. Distribution of Average Quality and Empathy Ratings for Chatbot and Physician Responses to Patient Questions



# How to use the unexpected, little understood effects of scaling LLMs to build better products

- **Chain-of-thought prompting:** A model prompted to explain itself performs orders of magnitude better than one that is not prompted to do so:

(Wei, J., Wang, X., Schuurmans, D., Bosma, M., Chi, E., Le, Q. and Zhou, D., 2022. Chain of thought prompting elicits reasoning in large language models. *arXiv preprint arXiv:2201.11903.*)

- LLMs of extraordinary size develop **sudden unexpected social bias:**

(Parrish, A., Chen, A., Nangia, N., Padmakumar, V., Phang, J., Thompson, J., Htut, P.M. and Bowman, S.R., 2021. BBQ: A hand-built bias benchmark for question answering. *arXiv preprint arXiv:2110.08193.*)

≡ SEARCH

FORTUNE

Subscribe Now

SIGN IN

Home News Tech Finance Leadership Well Recommends Fortune 500

TECH • A.I.

## Over just a few months, ChatGPT went from correctly answering a simple math problem 98% of the time to just 2%, study finds

BY PAOLO CONFINO

July 20, 2023 at 9:29 AM GMT+10



# How to use the unexpected, little understood effects of scaling LLMs to build better products

- Theory of mind: the ability to attribute to other people mental states different from our own. (models of the world, 'Sally-Anne test'):

(Kosinski, M., 2023. Theory of mind may have spontaneously emerged in large language models. *arXiv preprint arXiv:2302.02083*.)

- Moral self-correction: RLHF'ed models (Reinforcement Learning with Human Feedback) once reaching 22 Billion parameters, learn complex normative concepts of harm such as stereotyping, bias and discrimination and can be instructed to counteract them:

(Ganguli, D. et al., 2023. The Capacity for Moral Self-Correction in Large Language Models. *arXiv preprint arXiv:2302.07459*.)



# More ideas from the crossroads of neuroscience and AI

**MOTHERBOARD**  
TECH BY VICE

## Microsoft Now Claims GPT-4 Shows 'Sparks' of General Intelligence

The eyebrow-raising claim from Microsoft—which is banking on GPT putting it ahead of Google—contrasts with the model's clear limitations.

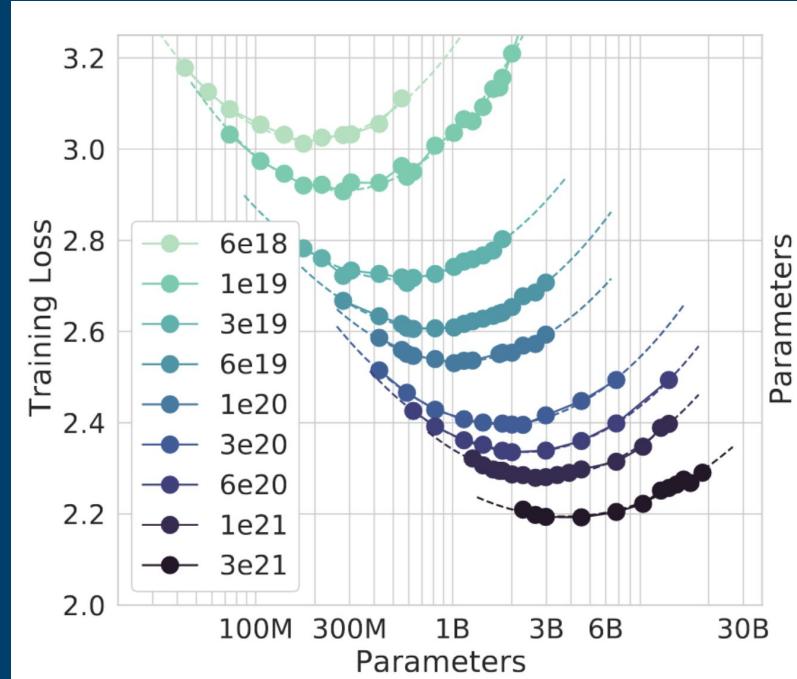


By [Chloe Xiang](#)

25 March 2023, 6:00am  [Share](#)  [Tweet](#)  [Snap](#)

Bubeck, S. et al., 2023. Sparks of artificial general intelligence: Early experiments with gpt-4. *arXiv preprint arXiv:2303.12712*.

# We might have seen the largest large language model for a while: scaling laws were misunderstood until recently



- constant compute (isoFLOP curves)
- **left of minima:** models are too small – training a larger model with less data would yield better model performance
- **right of minima:** models are too big – training a smaller model with more data would yield better model performance
- **scaling law:** for every increase in compute, data size and model size needs to be increased by same amount, i.e.: if compute increases by 10x, data and model size should increase by ~3.1x each.

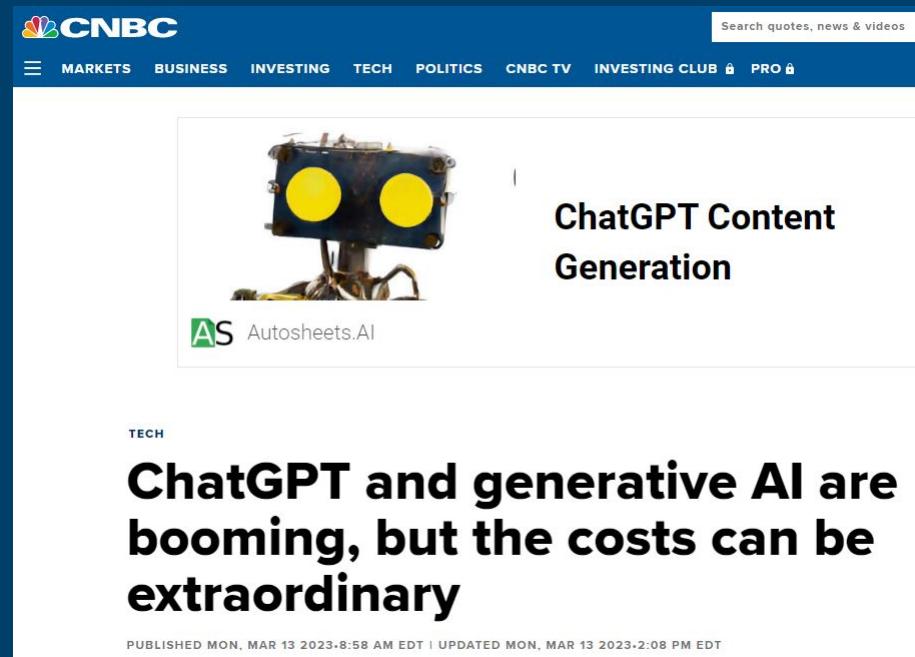
Hoffmann, J., Borgeaud, S., Mensch, A., Buchatskaya, E., Cai, T., Rutherford, E., Casas, D.D.L., Hendricks, L.A., Welbl, J., Clark, A. and Hennigan, T., 2022. Training compute-optimal large language models. *arXiv preprint arXiv:2203.15556*.

# LLM product development means making smart choices about data and models

Model	Size (# Parameters)	Training Tokens
LaMDA ( <a href="#">Thoppilan et al., 2022</a> )	137 Billion	168 Billion
GPT-3 ( <a href="#">Brown et al., 2020</a> )	175 Billion	300 Billion
Jurassic ( <a href="#">Lieber et al., 2021</a> )	178 Billion	300 Billion
Gopher ( <a href="#">Rae et al., 2021</a> )	280 Billion	300 Billion
MT-NLG 530B ( <a href="#">Smith et al., 2022</a> )	530 Billion	270 Billion

Parameters	FLOPs	FLOPs (in Gopher unit)	Tokens
400 Million	1.92e+19	1/29, 968	8.0 Billion
1 Billion	1.21e+20	1/4, 761	20.2 Billion
10 Billion	1.23e+22	1/46	205.1 Billion
67 Billion	5.76e+23	1	1.5 Trillion
175 Billion	3.85e+24	6.7	3.7 Trillion
280 Billion	9.90e+24	17.2	5.9 Trillion
520 Billion	3.43e+25	59.5	11.0 Trillion
1 Trillion	1.27e+26	221.3	21.2 Trillion
10 Trillion	1.30e+28	22515.9	216.2 Trillion

Hoffmann, J., Borgeaud, S., Mensch, A., Buchatskaya, E., Cai, T., Rutherford, E., Casas, D.D.L., Hendricks, L.A., Welbl, J., Clark, A. and Hennigan, T., 2022. Training compute-optimal large language models. *arXiv preprint arXiv:2203.15556*.



The screenshot shows a news article from CNBC. The header includes the NBC logo and the word "CNBC". Below the header is a navigation bar with links to MARKETS, BUSINESS, INVESTING, TECH, POLITICS, CNBC TV, INVESTING CLUB, and PRO. A search bar is located at the top right. The main content features a photograph of a robotic head with two large yellow circular eyes. To the right of the image, the text "ChatGPT Content Generation" is displayed in bold black font. Below the image, there is a logo for "AS Autosheets.AI" consisting of a green square with a white letter "A" and the text "Autosheets.AI". The article is categorized under the "TECH" section. The title of the article is "ChatGPT and generative AI are booming, but the costs can be extraordinary". At the bottom of the article, it says "PUBLISHED MON, MAR 13 2023 8:58 AM EDT | UPDATED MON, MAR 13 2023 2:08 PM EDT".

# Regulators are moving – but AI technology and adoption are moving faster

1. **Legal precedence** is currently being established through ongoing lawsuits

**What to watch out for:** how will AI/tech providers navigate the 'Section 230 vs. Copyright Law' conundrum?

2. **Governance frameworks** for the ethical design, development and use of AI are being created and established by regulatory and policy makers around the globe. (FDA, ISO, OECD, WHO, EU)

**What to watch out for:** how will the ratification of the EU AI Act impact global policy making, market dynamics and AI innovation?

MIT Technology Review

TECH POLICY

An early guide to policymaking on generative AI

How lawmakers are thinking about the risks of the latest tech revolution

By Tate Ryan-Mosley

March 27, 2023

Events Video Special Issues

VentureBeat

Search Subscribe GamesBeat

Artificial Intelligence ▾ Security ▾ Data Infrastructure ▾ Automation ▾ Enterprise Analytics ▾ More ▾

EU lawmakers pass draft of AI Act, includes copyright rules for generative AI

Sharon Goldman  
@sharongoldman

April 27, 2023 12:00 PM

f t in



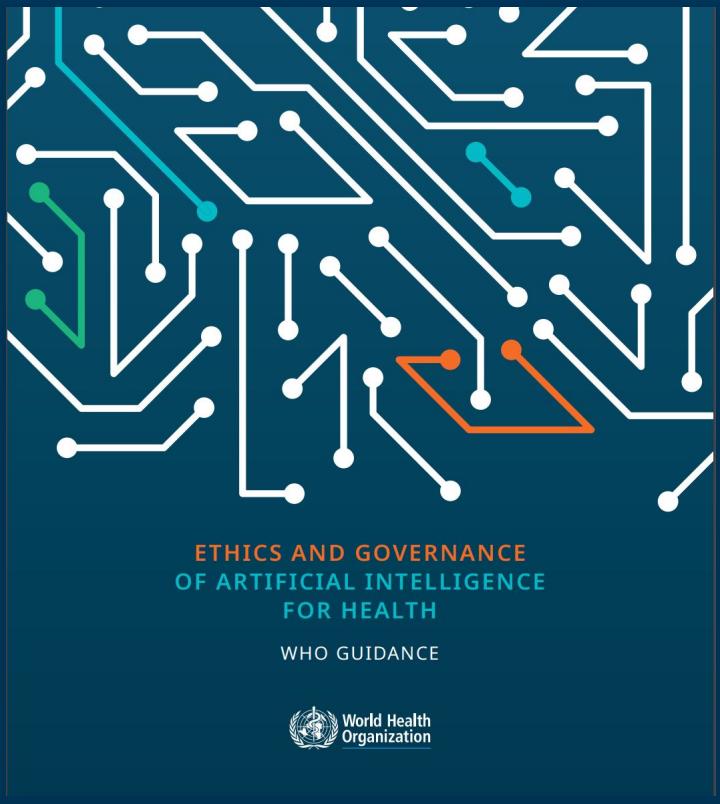
# Choosing the best LLM for an application – a moving goalpost problem

## Grading Foundation Model Providers' Compliance with the Draft EU AI Act

Source: Stanford Research on Foundation Models (CRFM), Institute for Human-Centered Artificial Intelligence (HAI)

	OpenAI	cohere	stability.ai	ANTHROPIC	Google	BigScience	Meta	AI21labs	ALEPH ALPHA	EleutherAI	
Draft AI Act Requirements	GPT-4	Cohere Command	Stable Diffusion v2	Claude	PaLM 2	BLOOM	LLaMA	Jurassic-2	Luminous	GPT-NeoX	Totals
Data sources	● ○ ○ ○	● ● ● ○	● ● ● ●	○ ○ ○ ○	● ● ○ ○	● ● ● ●	● ● ● ●	○ ○ ○ ○	○ ○ ○ ○	● ● ● ●	22
Data governance	● ● ○ ○	● ● ● ○	● ● ○ ○	○ ○ ○ ○	● ● ● ○	● ● ● ●	● ● ○ ○	○ ○ ○ ○	○ ○ ○ ○	● ● ○ ○	19
Copyrighted data	○ ○ ○ ○	○ ○ ○ ○	○ ○ ○ ○	○ ○ ○ ○	○ ○ ○ ○	● ● ○ ○	○ ○ ○ ○	○ ○ ○ ○	○ ○ ○ ○	● ● ○ ○	7
Compute	○ ○ ○ ○	○ ○ ○ ○	● ● ● ●	○ ○ ○ ○	○ ○ ○ ○	● ● ● ●	● ● ● ●	○ ○ ○ ○	● ○ ○ ○	● ● ○ ○	17
Energy	○ ○ ○ ○	● ○ ○ ○	● ● ○ ○	○ ○ ○ ○	○ ○ ○ ○	● ● ○ ○	● ● ○ ○	○ ○ ○ ○	○ ○ ○ ○	● ● ○ ○	16
Capabilities & limitations	● ● ● ●	● ● ● ○	● ● ● ●	● ○ ○ ○	● ● ● ●	● ● ○ ○	● ● ○ ○	● ● ○ ○	● ○ ○ ○	● ● ○ ○	27
Risks & mitigations	● ● ● ○	● ● ○ ○	● ○ ○ ○	● ○ ○ ○	● ● ○ ○	● ○ ○ ○	● ○ ○ ○	● ○ ○ ○	○ ○ ○ ○	● ○ ○ ○	16
Evaluations	● ● ● ●	● ○ ○ ○	○ ○ ○ ○	○ ○ ○ ○	● ● ○ ○	● ● ○ ○	● ○ ○ ○	○ ○ ○ ○	● ○ ○ ○	● ○ ○ ○	15
Testing	● ● ● ○	● ● ○ ○	○ ○ ○ ○	○ ○ ○ ○	● ● ○ ○	● ● ○ ○	○ ○ ○ ○	● ○ ○ ○	○ ○ ○ ○	○ ○ ○ ○	10
Machine-generated content	● ● ● ○	● ● ○ ○	○ ○ ○ ○	● ● ○ ○	● ● ○ ○	● ● ○ ○	● ● ○ ○	● ● ○ ○	● ○ ○ ○	● ● ○ ○	21
Member states	● ● ○ ○	○ ○ ○ ○	○ ○ ○ ○	● ● ○ ○	● ● ○ ○	○ ○ ○ ○	○ ○ ○ ○	○ ○ ○ ○	● ○ ○ ○	○ ○ ○ ○	9
Downstream documentation	● ● ○ ○	● ● ● ●	● ● ● ●	○ ○ ○ ○	● ● ● ●	● ● ○ ○	● ○ ○ ○	○ ○ ○ ○	○ ○ ○ ○	● ● ○ ○	24
Totals	25 / 48	23 / 48	22 / 48	7 / 48	27 / 48	36 / 48	21 / 48	8 / 48	5 / 48	29 / 48	

# A seminal report by the World Health Organisation leads the way to responsible design and use of (generative) AI



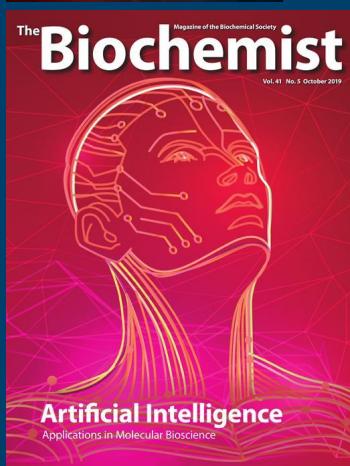
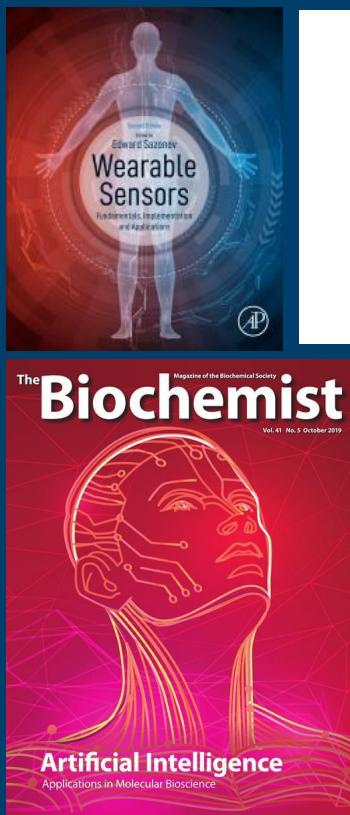
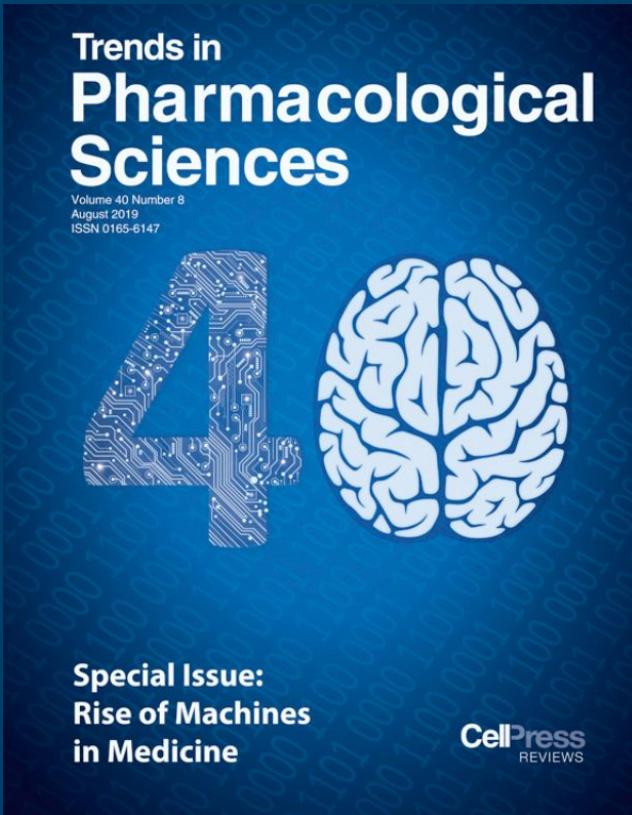
Design, development, and deployment framework for risk and benefit assessment of LLMs in health and medicine:

Harrer, S., 'Attention is not all you need: the complicated case of ethically using large language models in healthcare and medicine', **eBioMedicine (The Lancet)**, vol. 90, 104512, March 2023.

Implementation guide for responsible LLM-MLOps in health and medicine:

Harrer, S., 'Response to M. Trengove & coll regarding "Attention is not all you need: the complicated case of ethically using large language models in healthcare and medicine", **eBioMedicine (The Lancet)**, vol. 93, 104672, June 2023.

# Reading list – dig into the starters...



Trends in Pharmacological Sciences

Special Issue: Rise of Machines in Medicine

Review

## Artificial Intelligence for Clinical Trial Design

Stefan Harrer,<sup>1,\*</sup> Pratik Shah,<sup>2</sup> Bhavna Antony,<sup>1</sup> and Jianying Hu<sup>3</sup>

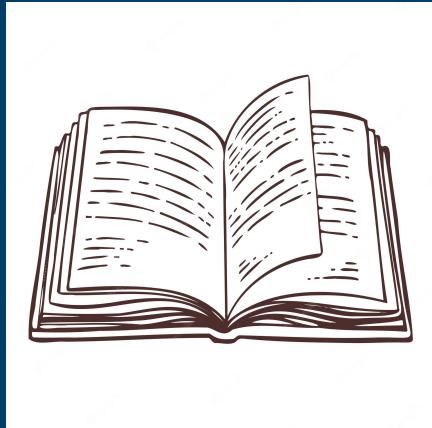
## Artificial Intelligence

# A new promising way for tackling the 'Pharma Dilemma': artificial intelligence for clinical trials

Stefan Harrer, Bhavna Antony, Akram Bayat and Jianying Hu (IBM Research, Australia, MIT Media Lab, USA and IBM T. J. Watson Research Center, USA)

Artificial intelligence (AI) is certainly not a panacea for solving the 'Pharma Dilemma,' in which the cost of producing new drugs continues to spiral. However, AI can be used to fundamentally change the way we perform essential steps in clinical trial design and execution, from cohort selection to patient monitoring. Merging AI and clinical expertise across engineering and medical disciplines to explore the impact of these changes on trial performance and success rates is one of the most promising leads we have for restoring efficiency and sustainability to the drug development cycle.

**...and watch out for the essential read: in bookstores  
starting September 2023**



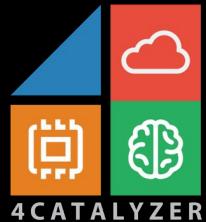
**“AI drives the digital transformation of pharma”, Harrer, Stefan (DHCRC), Menard, Jeffrey, Rivers, Michael (Roche Diagnostics), Green, Darren, Karpiak, Joel, Jeliazko, Jeliazkov, Shapovalov, Maxim, del Alamo, Diego, Sternke, Matt (GSK.ai), in *AI in Clinical Practice, prefaced by Eric Topol, 1<sup>st</sup> Edition*, in press (2023), Elsevier.**

[harrerbua@gmail.com](mailto:harrerbua@gmail.com)

<https://www.linkedin.com/in/stefanharrer/>

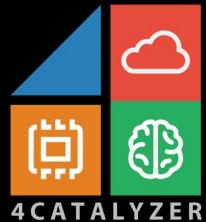
# AI x 4Catalyzer Summary Talk + Hackathon Topics

Andrew Gao  
Akshita Panigrahi  
July 28, 2023



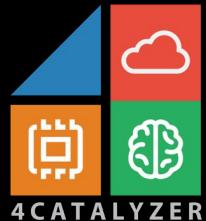
The goal of the LLMs for Bio hackathon  
is to:

Leverage **LLMs and Generative AI**  
towards a **key need or focus of**  
**4Catalyzer companies.**



Judges are looking for:

- Usefulness and relevance to 4Catalyzer
- Technological complexity/advancement
- Innovativeness
- Uniqueness
- **Integration of an LLM**



Judges do **not** want to see:

- Kaggle-style data science/ML projects
- Non-unique projects like Skin Cancer Detection using HAM1000 dataset
- Projects irrelevant to 4C
- Projects only using prompt engineering + the “raw” GPT-X API

# Broad Project Ideas:

- **Image-to-text:** Given a patient's MRI, retinal, or ultrasound scan image, the LLM will interpret the results, suggest next steps, potential health conditions, and suggestions.
  - This should go *beyond* merely classifying a single disease condition.
- **Sequence-to-text:** Given a patient's DNA sequence for a specific gene/region, the LLM will interpret the sequence. Example: "Based on your DNA sequence, I anticipate that you have blue eyes". "Based on your DNA sequence, I anticipate that you are at higher risk for skin cancer because....."
  - **Integrate** knowledge from databases like GeneCards, NCBI, OMIM, etc.
- Creating a platform or tool that helps researchers working on Directed Evolution for PETase and TdT enzymes

# 4C Overview - Abridged

# 4Catalyzer



4Catalyzer is a technology incubator founded by Dr. Jonathan Rothberg with the **mission to save lives and maximize societal impact.**

We work at the intersection of engineering, machine learning and natural sciences to support innovation in healthcare, life sciences, and climate tech.

# 4Catalyzer



HEALTHCARE



**HYPERFINE**



**Detect**



LIFE SCIENCE  
TOOLS



CLIMATE TECH



## CURRENT STATE - part I

**Health information is often inaccessible to many patients**



### **Confined to specialists' offices**

Imaging and testing is primarily confined to urban settings, large hospitals and specialists' offices, making access to care inconvenient and difficult for patients.



### **Large and complex devices, complex workflows**

Existing devices are immobile or complex and don't allow providers to meet patients where they are

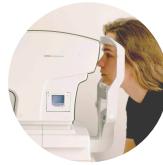


### **Prohibitively expensive**

The price point of existing devices puts them out of reach for most non-specialized facilities and use cases

## CURRENT STATE - part II

Life science tools are often inaccessible to many researchers



### Confined to specialized labs

testing is primarily confined to specialized labs and



### Large and complex devices

Existing devices are complex, have large physical overhead



### Prohibitively expensive

The price point of life sciences tools such as sequencing platforms and mass spectrometers (including reagents and test kits and data analytics) puts them out of reach for most non-specialized facilities and use cases

### CURRENT STATE - part III

Drug Development is costly and Drugs can be inaccessible to patients

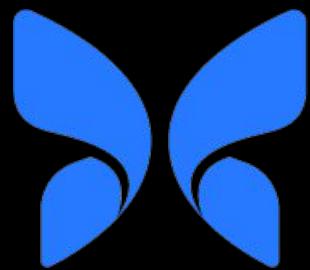
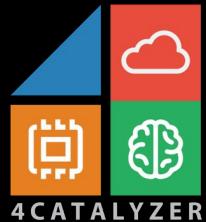
**Lengthy, complex, and costly** – high degree of uncertainty that a drug will actually succeed.

**Target Identification is challenging** – unknown pathophysiology for many disorders

**Heterogeneity of the patient population**  
-requires processing of large amounts of data (clinical phenotyping and subtyping)

**Lack of tools** for improved target identification and validation

# Hackathon Topics



# Butterfly™

Butterfly is commercializing the world's first ultrasound on a semiconductor chip. "BFLY" went public in February of 2021 on the NYSE raising half a billion dollars to continue to increase hospital connectivity as well as to further develop AI applications, ultimately enabling home use.

**NYSE: BFLY**

# Democratizing access to Ultrasound Imaging

- World's first full body ultrasound scanner on a semiconductor chip, fused with AI and cloud technology
- The quality of a \$60k device for a fraction of the price
- Two thirds of the world does not have access to medical imaging, and two thirds of diagnostic issues can be resolved with simple imaging.
- Saving lives on all 7 continents, partnership with the Gates Foundation for large scale deployments in Limited Resource Settings



# Butterfly Project Ideas

## Ultrasound Physician Assistant

An **LLM-based assistant that will use information immediately available during a doctor visit (e.g. primary care visit or checkup) to suggest an ultrasound imaging procedure appropriate for improving the overall assessment.** The LLM can be trained/fine tuned on published literature about the utility of ultrasound for assessing various conditions coupled with general medical knowledge linking information available during a doctor visit to target possible conditions. The resulting solution can take baseline information on the patient, augmented with additional information obtained during the visit (and possibly gleaned from a transcription of dialog between patient and doctor) and provide a real time suggestion on possible ultrasound scans that should be performed during the visit. The solution can also suggest training resources for obtaining the scan (from the public domain and from Butterfly Academy resources).

# Butterfly Project Idea (cont.)

Information utilized can include all or a subset of:

- A description (written or transcribed) of observations obtained in the course of the visit, including
  - Patient supplied descriptions of symptoms
  - Doctor's immediate observations
  - Live transcription of dialog between patient and doctor
- Basic information immediately available in the patient's EHR, such as age, sex, height, and ethnicity
- Metrics obtained from the current visit
  - Blood pressure
  - Heart rate
  - Weight/BMI
  - Blood oxygenation
- Historical health information including
  - Historical values of basic metrics, including weight, blood pressure, etc.
  - Lab test results
  - Past diagnoses
  - Current and past prescriptions
  - Write-ups from previous visits

**Follow on:** Utilizing the above information *and* an ultrasound image scan, suggest a diagnosis.

# Butterfly Project Ideas

## Ultrasound Pathology Simulator

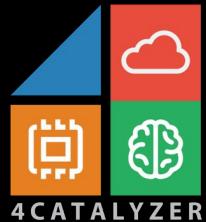
**Problem:** A challenge in advancing the adoption of point of care ultrasound (POCUS) is rooted in the **lack of training or knowledge in how to use ultrasound** during the course of practice. One benefit of handheld ultrasound is the ability to practice on one's self or on colleagues to gain proficiency in image acquisition. However, a key part of training is learning how to identify anomalies and pathologies in the images acquired by the operator and one barrier to training is the lack of access to patients with said pathologies.

**Idea:** Since it is quite easy to practice POCUS with a handheld, the ability to **simulate and introduce pathologies into a scan** would greatly increase the operator's confidence in their ability to leverage ultrasound to formulate a diagnosis. Recent capabilities in generative AI can potentially bridge this gap by **taking a live ultrasound image, a textual description of the desired pathology (i.e. "gallstones", "malignant breast tumor", "abdominal aortic aneurysm") and produce a new image** substantially similar to the given image, but with the requested pathology present

# Butterfly Project Ideas

## Ultrasound Radiologist in a Box

Build upon recent advances in semantic image AI to implement an anatomical feature and pathology detector for ultrasound images. Given an ultrasound image, **highlight key anatomical landmarks** (e.g. Iliac Vein, Right Kidney, gallbladder, etc.) and also **highlight abnormalities and potential pathologies** (e.g. gallstone, enlarged cardiac septum, etc.). Leverage published **literature** to extract images, image callouts, and captions for training. In addition to a textual **description** of observations, spatially indicate the **location** of various semantic components in the image.



# HYPERFINE

"HYPR" went public in 2021 on NASDAQ raising quarter billion dollars to advance the product pipeline, continue to increase hospital connectivity as well as to further develop AI to give clear images

**NASDAQ: HYPR**

# Democratizing access to MRI

- MRI is one of the safest forms of medical imaging - but is unavailable in 90% of the world - Hyperfine is changing that
- World's first Portable MRI
  - Low Cost, easy to transport to the patient's bedside, plugs into a wall outlet
- Powered by AI to give clear images and insights
- In use at over 100 sites worldwide



# Hyperfine Project Ideas

Learn to **generate the appearance of stroke lesions in and then insert them into another data**. This would help to train and improve stroke detection / segmentation projects. We will not have large-enough dataset for this yet (have to wait for ACTION-PMR), but there are some publicly available datasets this could be done with.

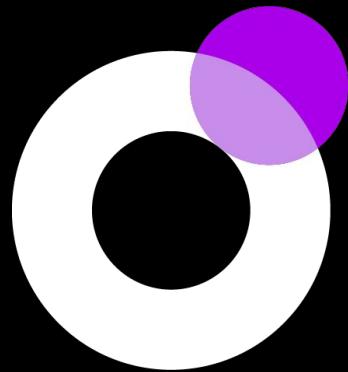
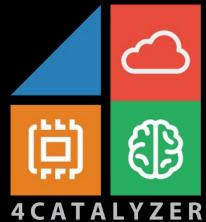
1. Use labeled dataset for **lesion simulation** and generation: <http://www.isles-challenge.org/>
2. **Insert** the **lesion** into another dataset, e.g.:  
<https://www.humanconnectome.org/study/hcp-young-adult/document/1200-subjects-data-release>

Extra points:

1. Retrain segmentation algorithm on generated lesion and submit to the challenge (this should improve ranking).
2. Downgrade the data by adding noise and reducing resolution as a simple transformation to low-field-like appearance.

# Hyperfine Project Ideas

1. Use LLM to train (or finetune opensource LLM model from Meta) on large scale radiology report dataset (e.g.: <https://physionet.org/content/mimic-cxr/2.0.0/>). I wish we had one for MRI.
2. Develop system to retrieve reports with specific queries about the findings. This will be much more powerful than query search.



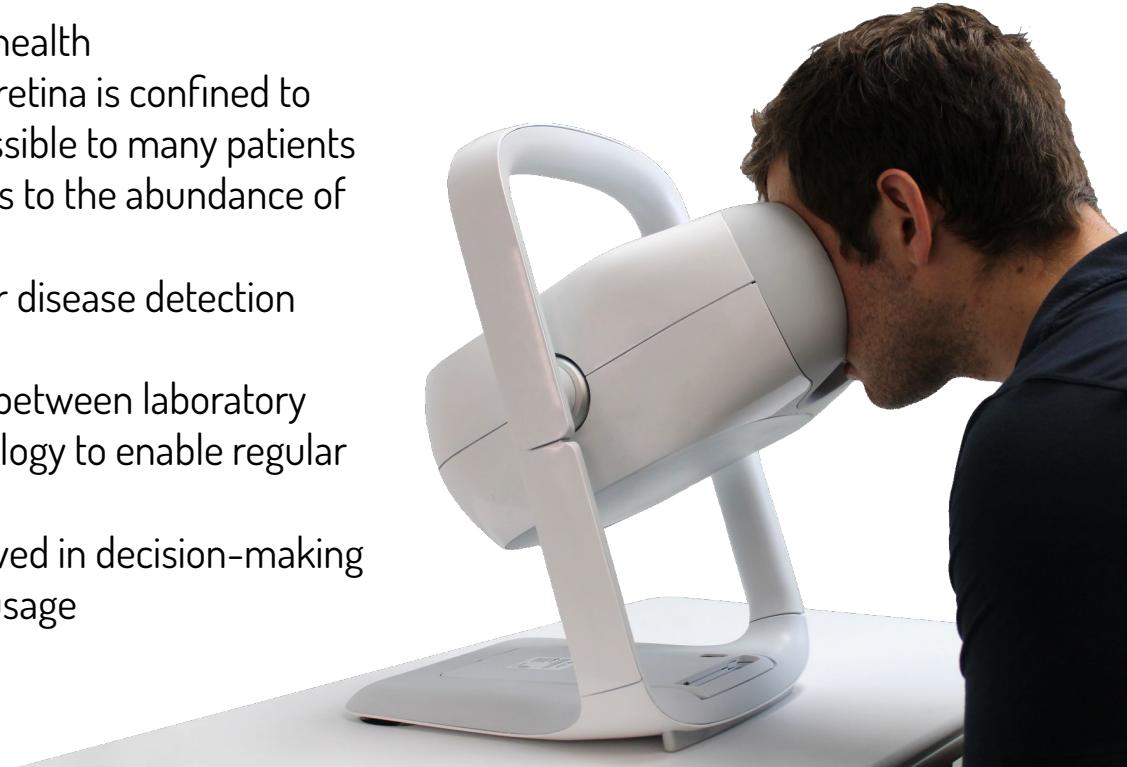
# identifeye

[HEALTH]

identifeye HEALTH (fka Tesseract Health) raised \$80M through Series B to develop and launch its first product. We are building an intuitive, consumer friendly, medical device to capture health information from the eye.

# Create a New Branch in Diagnostics - Making Retinal Imaging as simple as an eye selfie

- Eyes are a window to the body and health
- Access to health information in the retina is confined to specialists' offices and often inaccessible to many patients
- Our mission is to democratize access to the abundance of health information in the eye
- Prevent vision loss by making ocular disease detection easier and more accessible
- Create a new branch in diagnostics between laboratory medicine (e.g. bloodwork) and radiology to enable regular non-invasive health monitoring
- Empower patients to be more involved in decision-making by reducing barriers to access and usage



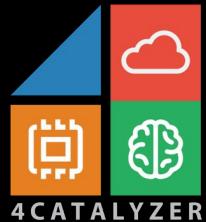
# Identifeye Project Ideas (open-ended)

We are currently focusing on fundus images - and our first application is Diabetic Retinopathy and in the pipeline we have a big focus on cardiovascular.

Some of the most straight forward from the public databases:

- MESSIDOR/MESSIDOR-2/IDRiD and Kaggle DR set for DR
- DRIVE and STARE for a variety but they're often used for vessel segmentation
- AIROGS and REFUGE for glaucoma
- AREDS for AMD
- UKB is great but access is not trivial to get access to

If you want something easy and straightforward to obtain, start with the **DR** and maybe **DRIVE** and **STARE**



# QuantumSi

Quantum-Si has built the world's first single molecule protein sequencer on a semiconductor chip. QSI went public in 2021 raising over \$500 million to revolutionize the future of diagnostics.

**NASDAQ: QSI**

# Single molecule protein sequencing

- First ever single molecule protein sequencing device
- DNA tells you what may happen, proteins tell you what is happening/about to happen
  - Could give us the ability to predict a heart attack before it happens with high accuracy
- Deeper proteomic insights will advance science and human health

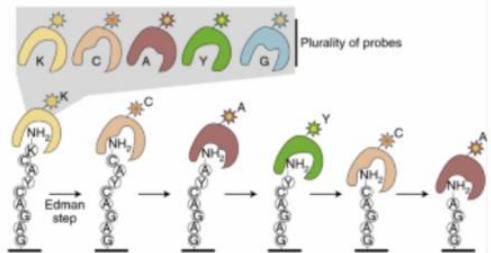


# Quantum-Si Project Ideas

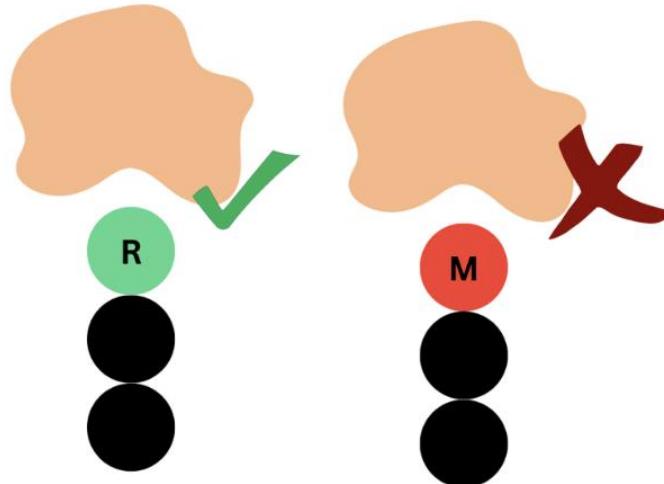
Amino acids to target: G, P, or M

Designing proteins that selectively bind to a specific amino acid from the side (N-terminus) and two more amino acids, of any residue.

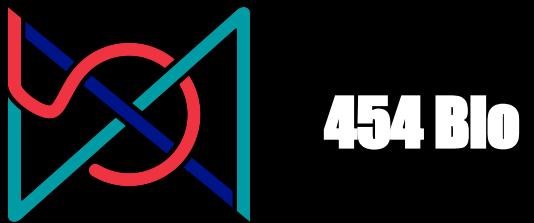
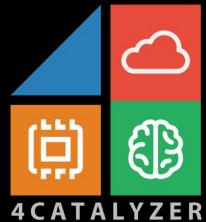
b Sequencing by N-terminal probes



Real-time dynamic single-molecule protein sequencing on an integrated semiconductor device  
(Reed et. al)  
<https://www.science.org/doi/10.1126/science.abo7651>



Example: this binder selectively binds to arginine, but not other amino acids



Founded in 2022, 454 Bio is building the first Next Generation DNA sequencer (NGS) fit for at-home use

# Decentralized On-site Sequencing

Today, personal genomics is limited by expensive and cumbersome equipment – turnaround time is slow, and sequencing requires expertise and resources

454 Bio's mission is to bring genomics out of the lab and enable universal access to affordable and fast DNA sequencing.

Low Cost device, easy to use kits, seamless data transfer for results without expertise

454 Bio is supporting a wide range of use cases in public health, research, and consumer genomics.



# 454 Bio Project Ideas

## Variant Evolution

Link SARS variant evolution with geography and/or clinical severity, and then try to predict which novel variants may be "next". For example, it could be any respiratory virus..

Find a link between sets of variants across a human genome to disease (any) with onset.

SARS datasets: <https://www.ncbi.nlm.nih.gov/sars-cov-2/>

## Cancer Genome Stability

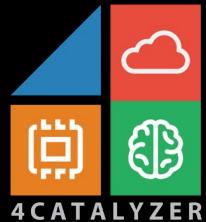
Identify regions in the human genome which are stable in cancer or the opposite.

Manually curated somatic cancer mutation data: <https://cancer.sanger.ac.uk/cosmic/download>

Cancer study data: <https://www.cbiportal.org/datasets>

Genomic Data Commons Data Portal: <https://portal.gdc.cancer.gov/>

Raw SNP data: <https://www.ncbi.nlm.nih.gov/snp/>



# PROTEIN EVOLUTION

Protein Evolution is leveraging recent breakthroughs in natural science and artificial intelligence to design enzymes to break down end-of-life textile and plastic waste into the building blocks that make up new textile and plastic products. Protein Evolution aims to help the chemicals industry transition to a lower-carbon, circular economy.

# Infinitely recyclable High quality plastic

- Identify the waste - water bottle, car tire, or piece of clothing.
- Engineer enzymes - Uniting natural science and artificial intelligence, develop enzymes that break down the waste source so we can recycle it in an economical, sustainable way.
- Break down the plastic into its “building blocks” in a low-emission, eco-friendly process.
- Reproduce materials by using these “building blocks” to create good-as-new plastic bottles, textiles, and other infinitely renewable plastic products.



# PEI Project Ideas

## AI Lab Trainer

Using public resources, train an AI model that can teach lab scientists how to perform a new experiment. The model should generate lab protocols and explain the caveats of the experiment and provide information about expected results.

We envision it to be a conversational AI model but please feel free to think outside the box.

## AI Safety Inspector

Given lab protocols, download all relevant MSDS sheets and generate a safety report. For example, "this experiment requires BSL-3" or "this experiment will accidentally create mustard gas as an intermediate". It would be great if the AI Safety Inspector could also suggest solutions to mitigate the safety issues. As an additional aim, AI Safety Inspector could highlight where relevant information is missing in the protocol and propose to add it. Relevant AutoGPT paper:

<https://arxiv.org/ftp/arxiv/papers/2304/2304.05332.pdf>. We envision this to be a Generative AI model but please feel free to think outside the box.

# PEI Project Ideas

## A protein-specific, scalable vector search platform

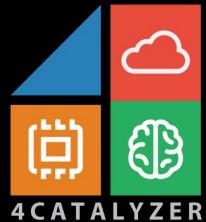
Given a protein sequence, find the nearest neighbors in function space. For example, given an enzyme, find other enzymes in the database that have similar optimum pH.

## ProteinLLM to predict optimum pH

Given an enzyme, predict the optimum pH at which the enzyme would perform.

This is a difficult problem. See this relevant paper:

<https://www.biorxiv.org/content/10.1101/2023.06.22.544776v1>. The RSME can be as high as 2-3 pH units, which is too high an error to be useful. It would be interesting to even learn more about why the error is so high and how we can improve it. We envision this to be a discriminative AI model but please feel free to think outside the box.



# Detect

In response to the global pandemic, Detect built and brought-to-market a PCR-quality rapid molecular home Covid-19 test authorized for EUA by the FDA. It is now focusing on a platform for POC and home use.

# Access to rapid PCR quality testing

- Detect's proprietary technology serves as the platform for its future home tests, including the next-generation Detect Covid+Flu Test, as well as rapid molecular home tests for respiratory health, Strep and STIs.
- Super easy to use, low cost, and fast
- Enables stakeholders to test on site, without sending to a lab

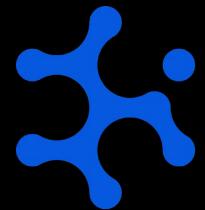
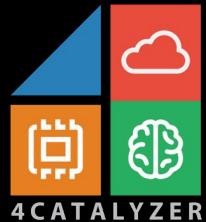








Detect

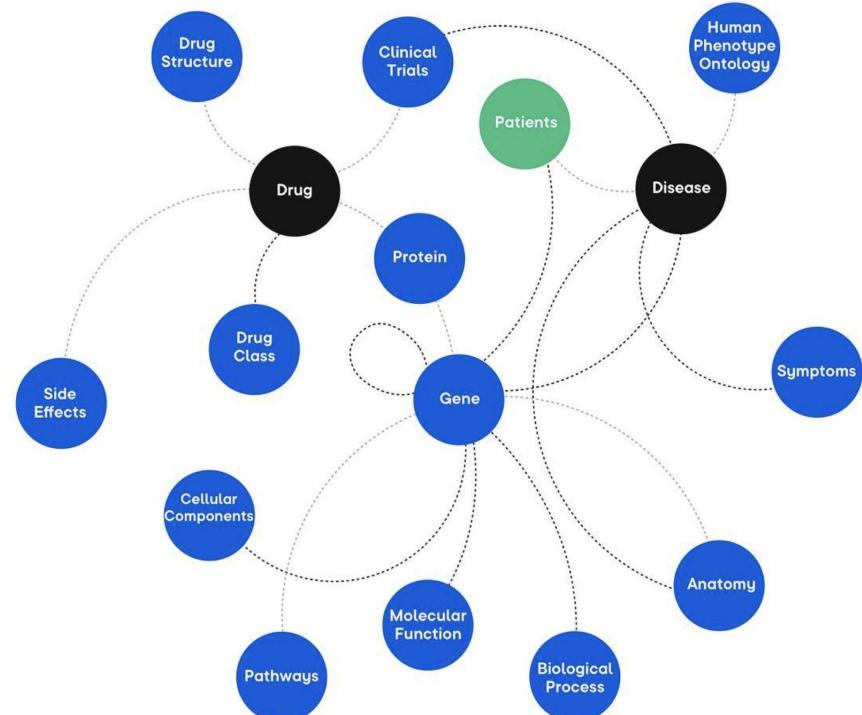


# AI Therapeutics

AI therapeutics is identifying and developing drugs  
& treatments for rare orphan conditions.

# Identifying and developing drugs for rare conditions

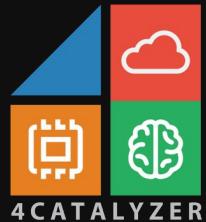
- Uses AI Algorithms to identify promising assets to add to their pipeline
- AI analyzes what these safe drugs can be used for
- Helping find treat for orphan/rare diseases
- Current clinical stage drug candidates target several orphan disorders including amyotrophic lateral sclerosis, pulmonary arterial hypertension, bronchiolitis obliterans, pulmonary sarcoidosis, Ewing sarcoma, rhabdoid tumor, and SWI/SNF mutated or dysregulated cancers.



# AI Therapeutics Project Ideas

Using gene expression information or pathway information on compounds that have been shown to be safe in man to find new indications.

Thousands of compounds have been in man and are safe but likely had the wrong indication.



# Liminal Sciences

Liminal Sciences is developing a non- invasive brain monitor to sense, understand, and ultimately heal the brain.

# Liminal is building the world's first non-invasive brain monitor

- The Brain is as important to monitor as the heart – but today there is no way to do this without drilling a hole in the patients skull
- Liminal is building the first wearable brain monitor for acute and chronic conditions including epilepsy, stroke, and traumatic brain injury.
- Enabling a brain monitor as ubiquitous as the heart monitor
- Partnered with AI for quicker and more valuable health insights.



# Liminal Sciences Project Ideas

Using brain ultrasound and AI for diagnosis of brain-related conditions, disorders, etc.

# Hackathon Logistics

# Guidelines

1. Please abide by posted signage
2. Do **not** enter any rooms or cabins inside the Gene Chaser.
3. **Use** the two bathrooms, the “Dayhead” by the Lab and the bathroom adjacent to the Gene and Tonic bar (inside a container).
4. **Meals** will be announced in the Discord when almost arrived, and when arrived. If you did not request a vegetarian meal beforehand, please do not take one.
5. **Snacks and beverages** (water, sodas, energy drinks, etc) are in the lounge upstairs.
6. All meals are **halal**.
7. Turn **on** your Discord notifications!

# Guidelines

1. Ask Andrew or Akshita during the event if you have questions. You may also ping us (@ us) in Discord.
2. Please be courteous with the shared Lambda compute.

# Professor Zhang Yang

Sunday at 2 PM.

Professor Zhang Yang will be giving a talk on protein structure prediction, AI, and related topics.

Professor Zhang is the creator of I-TASSER which is the State of the Art protein structure prediction algorithm on Earth (won the CASP 2022 competition).  
I-TASSER has won CASP eight out of the last nine times.

All teams will be expected to attend the seminar.

# On-Site Resources

Dr. Bonnie Gould Rothberg: (medical doctor, familiar with all 4C companies)

Jose: Expert at 454 (DNA Sequencing technology)

David Honeybun: Engineer at 4Catalyzer, familiar with all aspects of various 4C devices

Andrew Gao: AI intern at 4C, familiar with LLMs

Akshita Panigrahi: 454 Intern

Herschel Meadow: Engineer, familiar with Butterfly and Hyperfine

# On-Site Resources

Extension cords (Singapore compatible) (hangar)

Whiteboards and markers (hangar)

Snacks, cold drinks, coffee, etc. (lounge)

Meals will be placed in the lounge at the back counter.

# Project Submission

Projects will be strictly due at 2:00 PM on July 30th. Late submissions can not be considered out of fairness.

Your team will submit one Google Form answering questions about your project. Ensure that you allot time to write the questions. The Google Form link will be **shared** in the Discord on Sunday.

In the Google Form you will be asked to share a **live link** to a working demo **or** a screen recording.

Email a .zip file containing your code, data, and a well-written Readme. **Poorly written Readmes** without sufficient detail to replicate results or launch your program will be heavily **penalized**.

**You will still be allowed to work on your project during the demos after 2:00 PM. But do not work on your project while speakers are giving speeches.**

# Project Demos

As part of judging, you will be doing a demo of your project. You have **2** minutes to demo and **3** minutes for questions from judges. This will be strictly enforced to ensure that teams can present their awesome projects in a timely manner out of respect to hackers and judges.

Share your project demo Google Slideshow as a **working link** to  
[aqao@4catalyzer.com](mailto:aqao@4catalyzer.com)

Ensure that your Sharing Settings are properly configured. We may have to skip your demo if your slideshow is not presentable.

# Team Formation

Start thinking of a 30 second pitch:

Your background, your skills, what you are excited to work on, what you are looking for

Lambda builds the  
world's best deep  
learning public &  
private clouds

# About Lambda

Founded by ML Engineers with published research at top AI conferences



**Founded by ML Engineers**

NVIDIA's AI Partner of the Year for 2020, 2021, 2022 & 2023



**100% focused on Deep Learning**

**NVIDIA's Top AI Partner**

**Public & Private Cloud**



Everything we build serves ML Engineers & Researchers



Only company in the world that provides GPU compute from desktop to datacenter to cloud

# What Lambda builds

01

## Public Cloud

Only Public Cloud dedicated to Deep Learning workloads & distributed training.

02

## Private Cloud

Engineered high-performance clusters designed, built & supported by ML Engineers.

03

## Software

Plug and play AI stack that solves CUDA & PyTorch driver and dependency issues.

# Lambda Cloud Resources



[Cloud Datasheet](#)

[A10 Datasheet](#)

[A10 Product Page](#)



[Video](#)

On-Demand Cloud - How  
to Spin-Up GPU Instance  
for Cloud



[Video](#)

On-Demand Cloud - How  
to Sign-up, Spin-up GPU,  
Launch GenAI Demo



[Video](#)

On-Demand Cloud - How  
to Sign-Up for Cloud

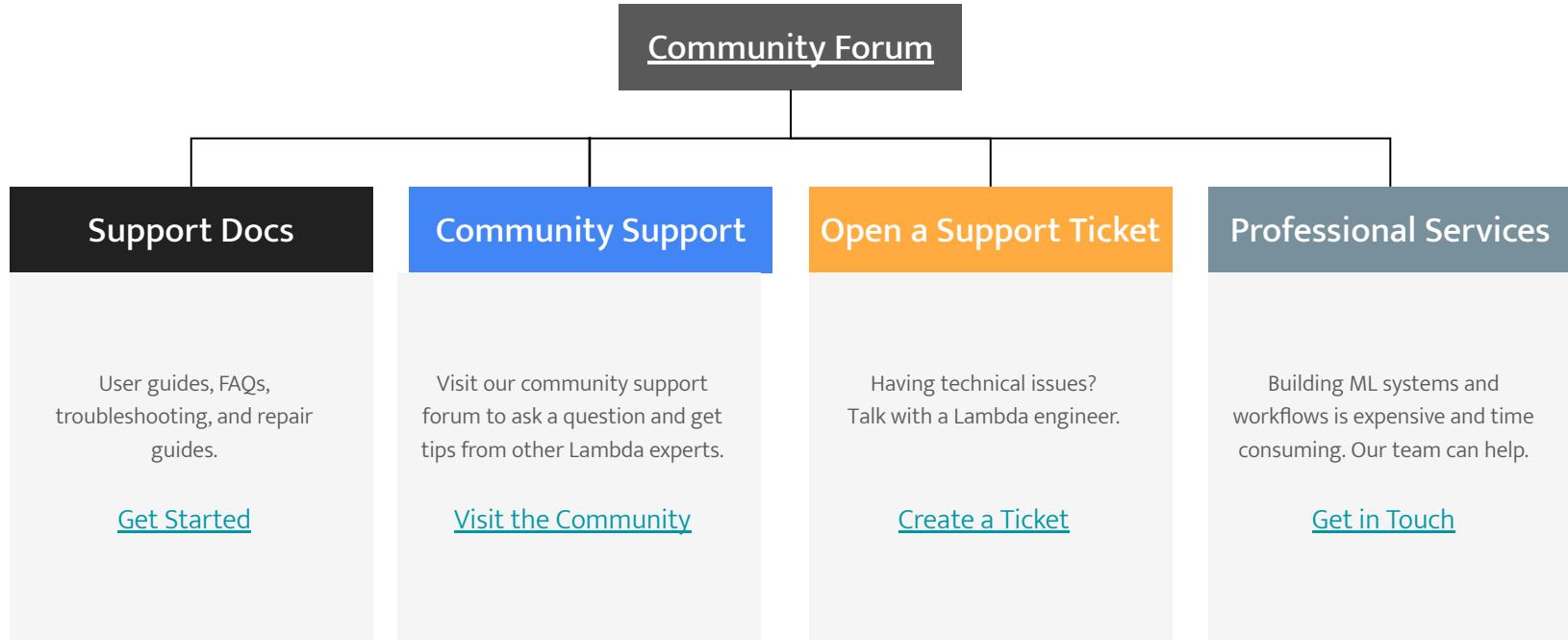


[Video](#)

On-Demand Cloud - How  
to Launch GenAI Demo

# Lambda Technical Support

Looking for quick answers? Reach out to the Lambda [Cloud team](#)



**Thank you! Let's have a great hackathon!**

If you are still looking for a team, don't disperse just yet!

# Team Mixer

Grab the mic and give a 30 second - 1 minute quick intro of yourself

1. Background/Experiences
2. What you want to work on
3. What are you looking for