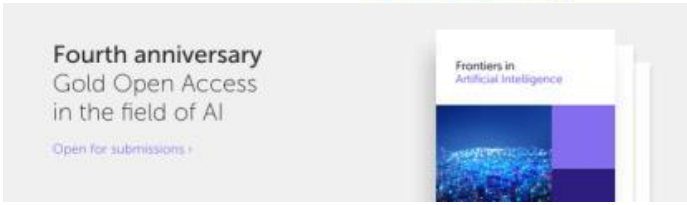
JOHNS HOPKINS
BLOOMBERG SCHOOL
of PUBLIC HEALTH

Johns Hopkins University
Center for Alternatives to Animal Testing



Thomas Hartung & team

Toward ToxAIcology

Slides available:

frontiers in Artificial Intelligence — **Field Chief Editor**

frontiers in Big Data — **Chief Editor Medicine & Public Health**

AstraZeneca

ATCC

AxoSim
Human Data, Faster.
VP, shareholder

MPS

inSphero

Apellis

Pyrogen

MERCK MILLIPORE

BIOMÉRIEUX

AI

UL **Consultant**

Green Chemistry Advisory Panel

ToxTrack
Consultant, shareholder
In preparation: Insilica LLC

Thanking our sponsors

Space for You!

Recent

Philantropic

…and individuals

Call for a Human Exposome Project, in press

Food for Thought ...

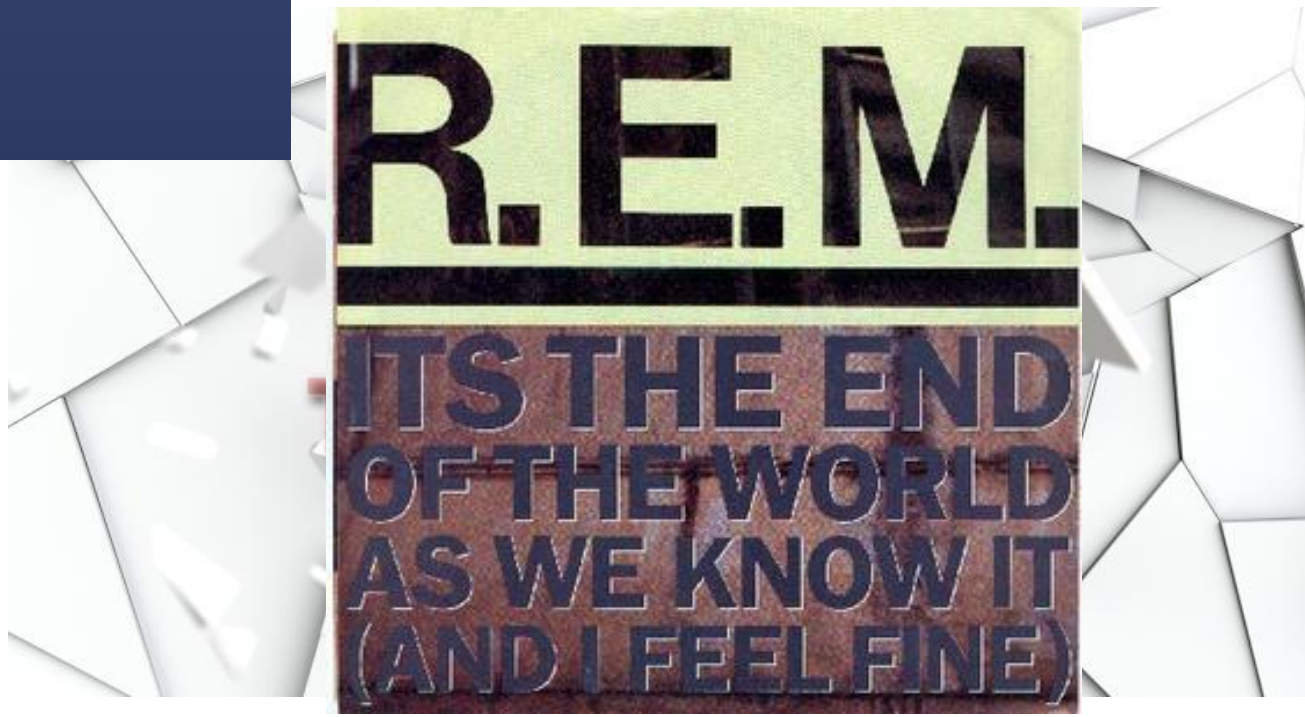
ALTEX 2023

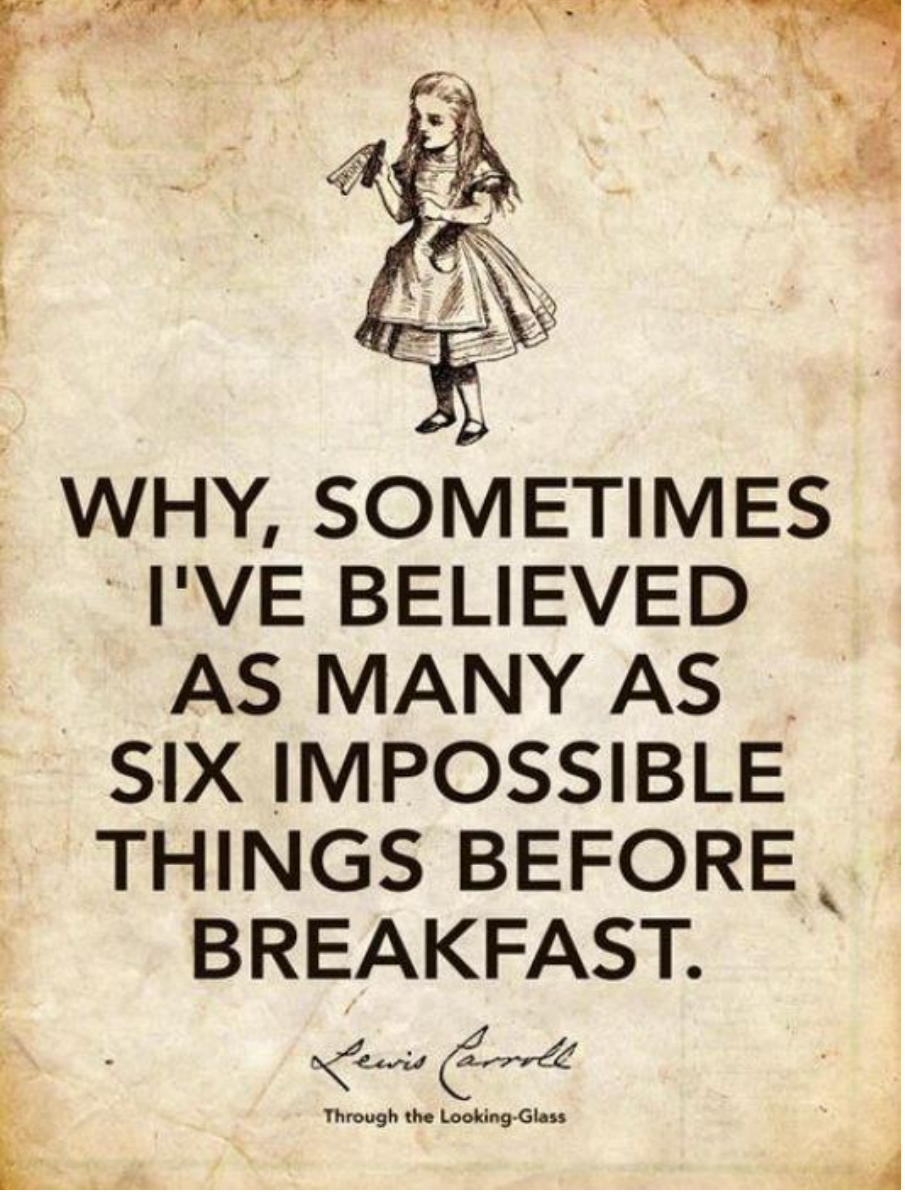
A Call for a Human Exposome Project

Thomas Hartung[1,2]

Data: +60% per year
= 90% in last two years
Computer: +40% per year
(Moore's law)
AI: +700% per year since 2010

Together increase >1 billion-fold over last 60 years





1. **There is a better way to play chess**
   2022 DeepMind: AlphaZero

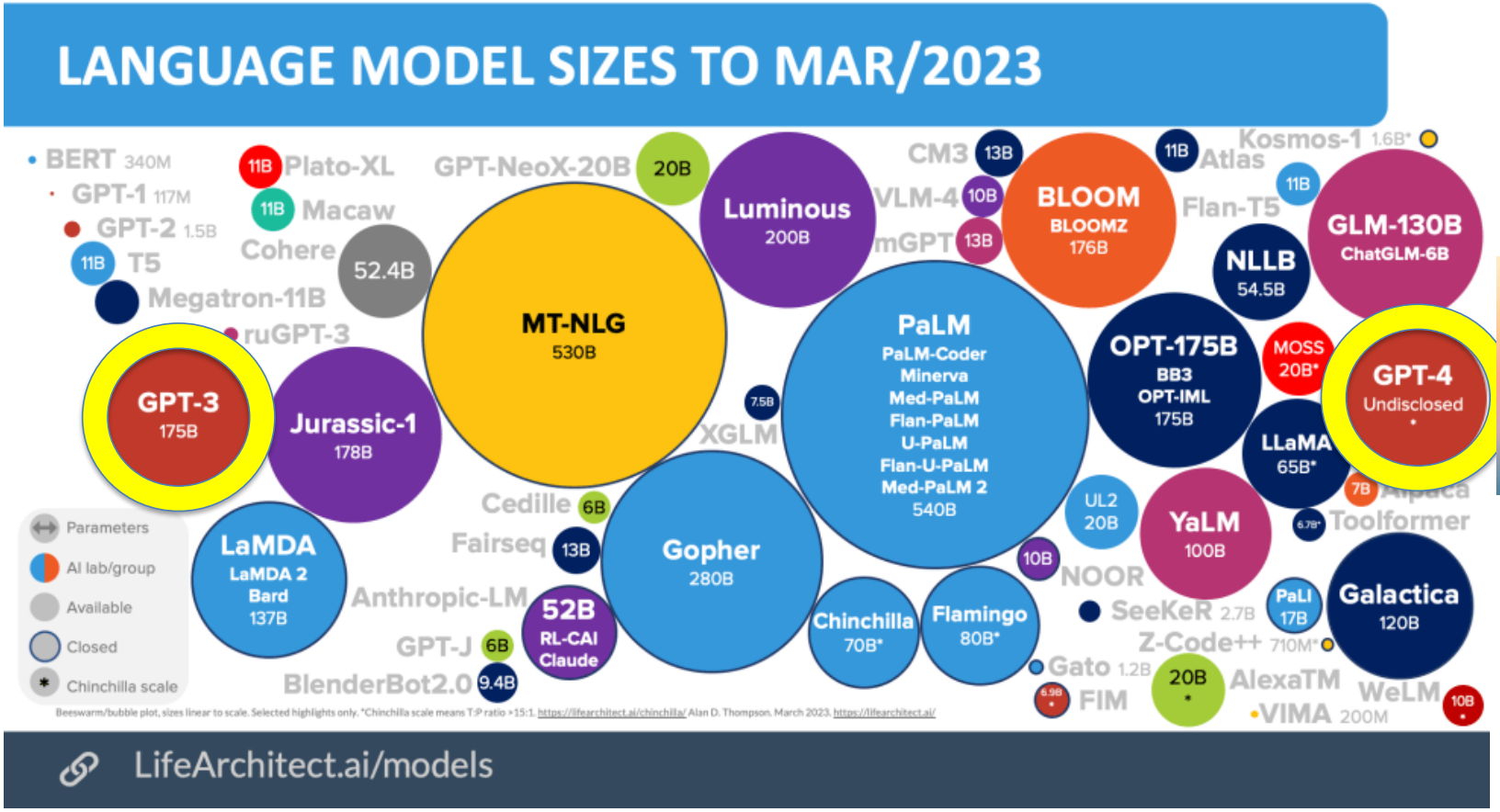
2. **The structure of all proteins can be predicted from gene sequence**
   2022 DeepMind: AlphaFold

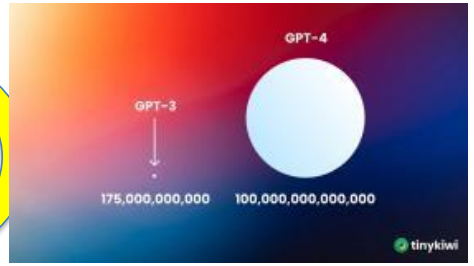
3. **A computer is better than (most) lawyers**
   2023 OpenAI: GPT-4

4. **A computer exceeds computational capacity of a human brain**
   2022 Frontier Computer exceeds 1 exaflop

5. **AI can design drugs**
   2022 – 18 AI-first drugs in clinical trials

6. **AI wins art contests**
   2022 Midjourney





Variables trained on

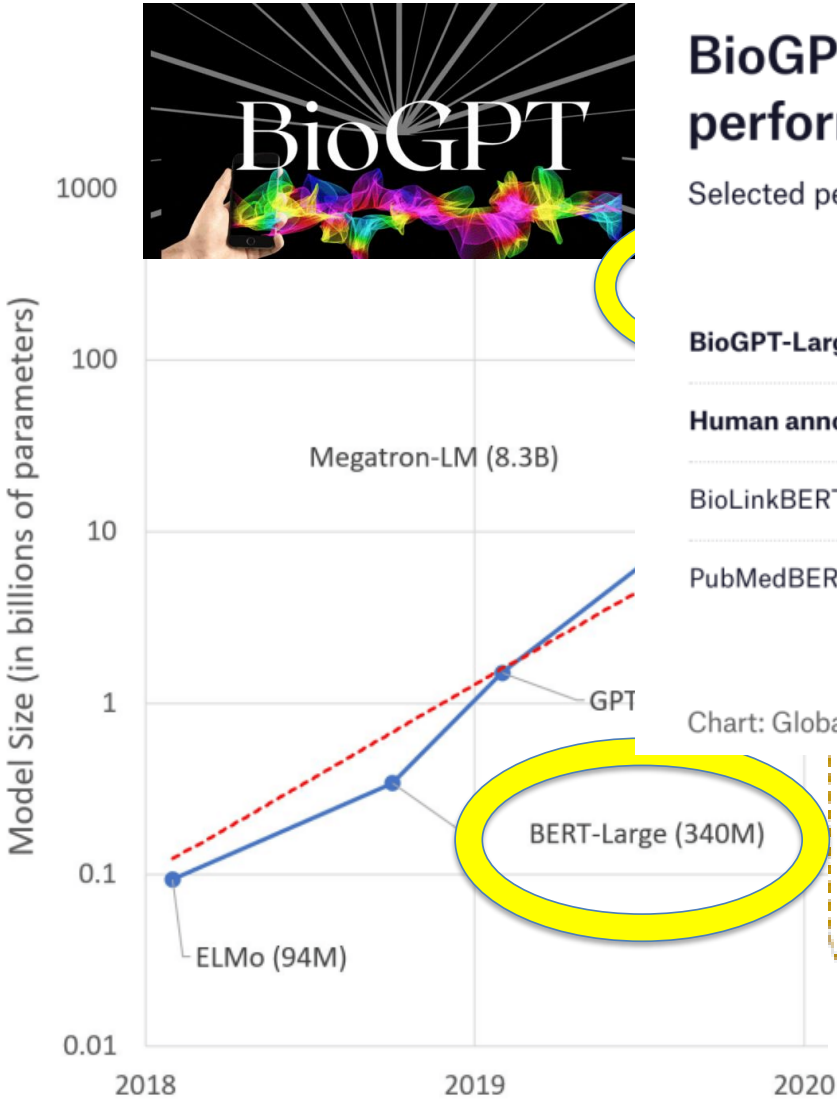
14 March 2023 Launch of GPT-4

https://lifearchitect.ai/gpt-4/

GPT-4 VS HUMAN TESTS (APR/2023)

Selected highlights only. Percentiles; 50 refers to the 50th percentile as average, and may not be the testing average for some tests. Alan D. Thompson. April 2023. https://lifearchitect.ai/iq-testing-ai/

LifeArchitect.ai/iq-testing-ai



BioGPT and human annotator have comparable performance in biomedical research test

Selected performances on PubMedQA, which tests biomedical language processing

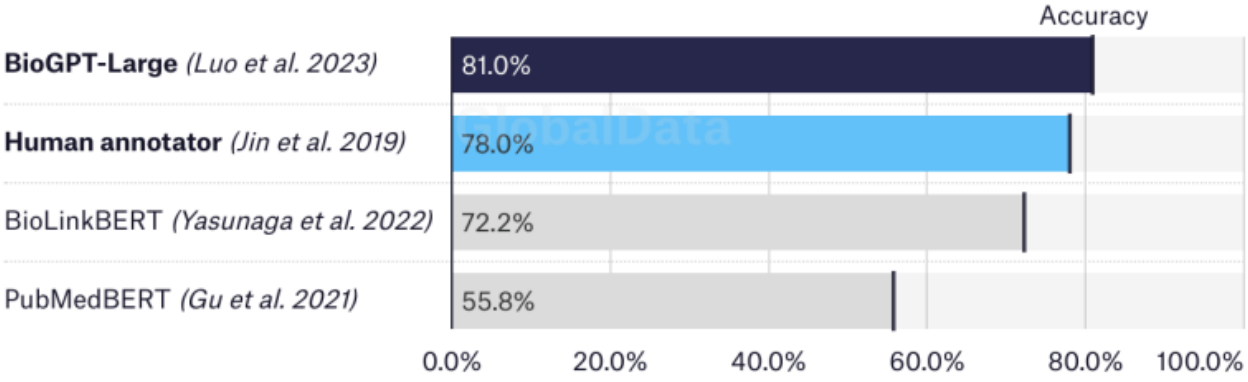
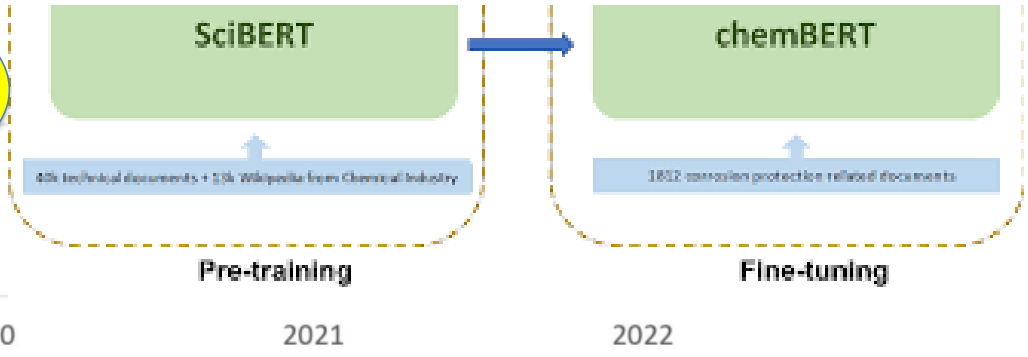


Chart: GlobalData • Source: PubMedQA



	PRE-2020	2020	2022	2023?	2025?	2030?
TEXT	Spam detection Translation Basic Q&A	Basic copy writing First drafts	Longer form Second drafts	Vertical fine tuning gets good (scientific papers, etc)	Final drafts better than the human average	Final drafts better than professional writers
CODE	1-line auto-complete	Multi-line generation	Longer form Better accuracy	More languages More verticals	Text to product (draft)	Text to product (final), better than full-time developers
IMAGES			Art Logos	Mock-ups (product design, architecture,	Final drafts (product design, architecture, etc.)	Final drafts better than professional artists, designers, photographers)
				Second drafts	AI Roblox Video games and movies are personalized dreams	

Almost there Ready for prime time

A personal take on science and society

Nature 20 Apr 2023

World view

Open generative AI models are a way forward for science

By Arthur Spirling

Researchers should stop using proprietary large language models and develop transparent ones to ensure reproducibility.

With open-source large

be the same, or even whether the technology will still be supported? GPT-3, released last November by OpenAI in San Francisco, California, has already been supplanted by GPT-4, and presumably supporting the older LLM will soon

ONTOX

EU ONTOX project ($20 million, 2021-2026) to address liver, kidney and developing brain

1. The largest toxicological database ever
2. Automatic paper selection and information extraction
3. Automatic extension of AOPs and Physiological Maps
4. Estimation of internal dose from exposure
5. Probability of hazard from chemical structure and perturbation of biology
6. Probability of risk enabling information economy

Six impossible things
I believe ONTOX will deliver

Internet Data

Literature

[redacted]

Big Data & A.I.

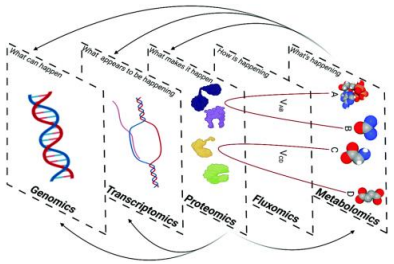
Database

Legacy Data

Robotized Testing

High Content Imaging

~omics Technoloogies

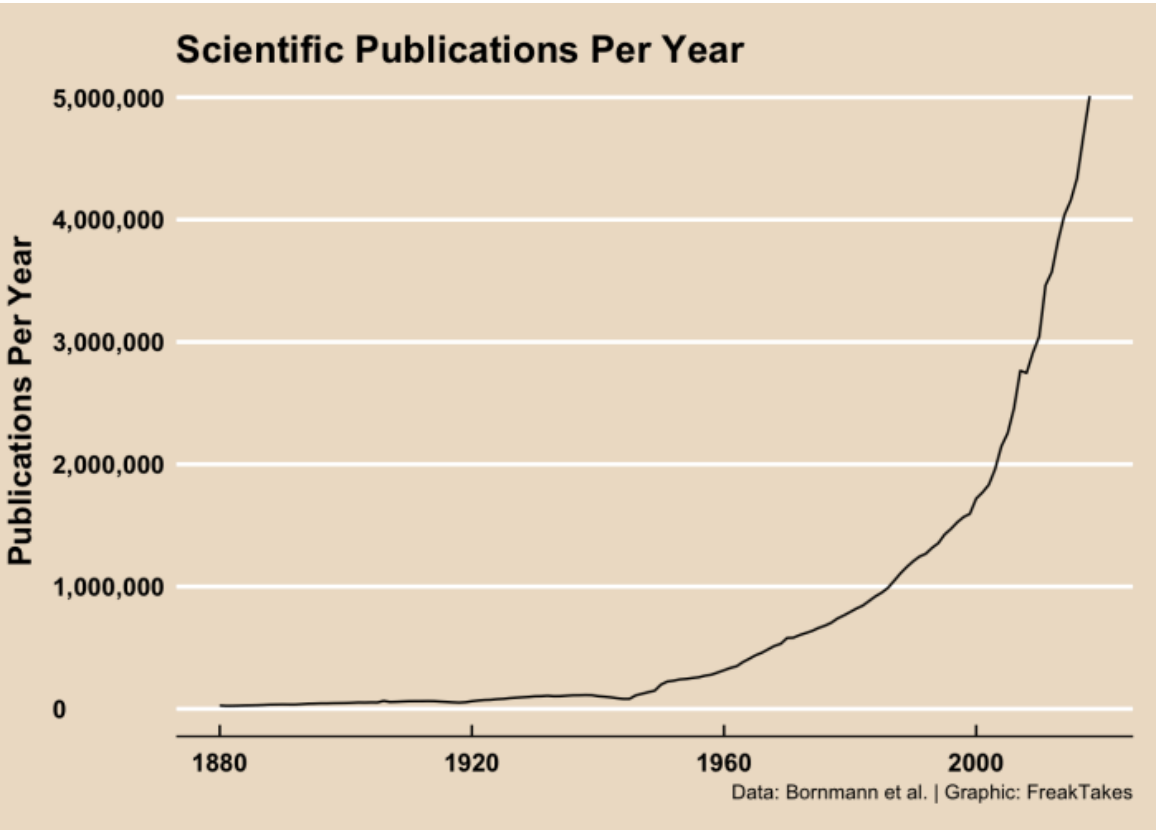


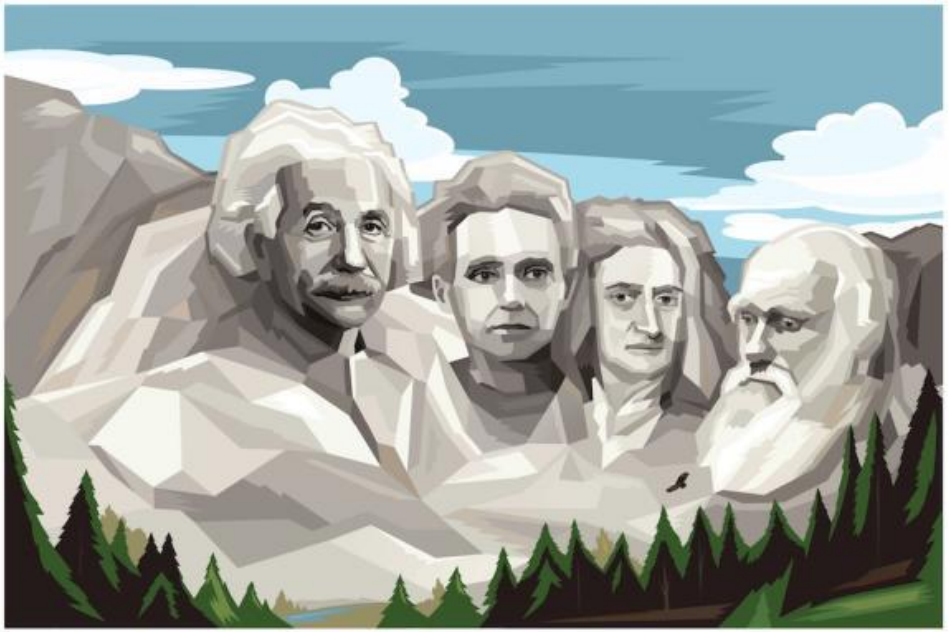
ONTOX

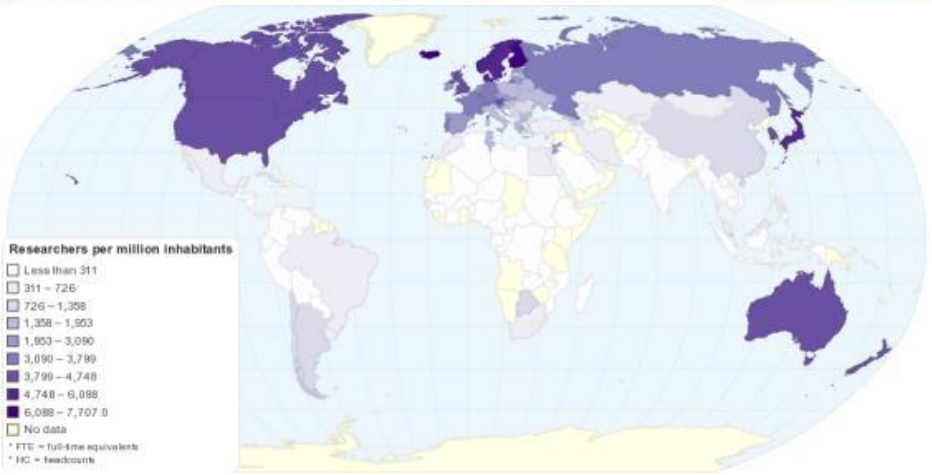
Tox21

ToxCast™

8.8 million researchers world-wide

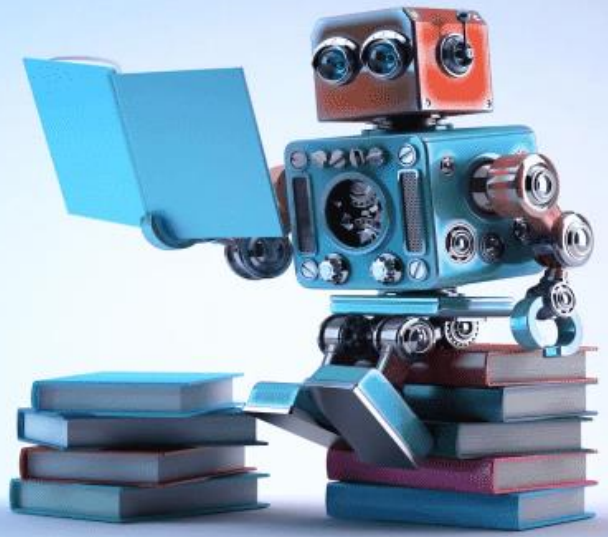




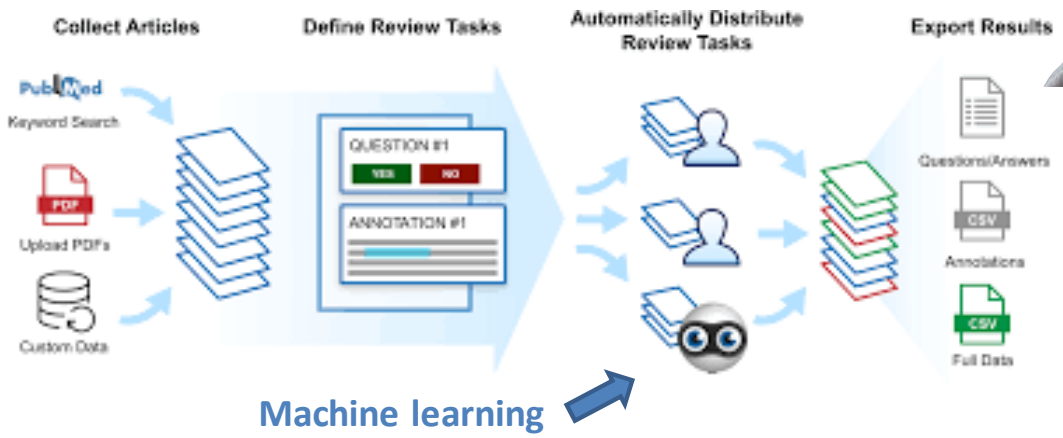


Researchers per million inhabitants

Data extraction from literature, reports & databases



ToxTrack



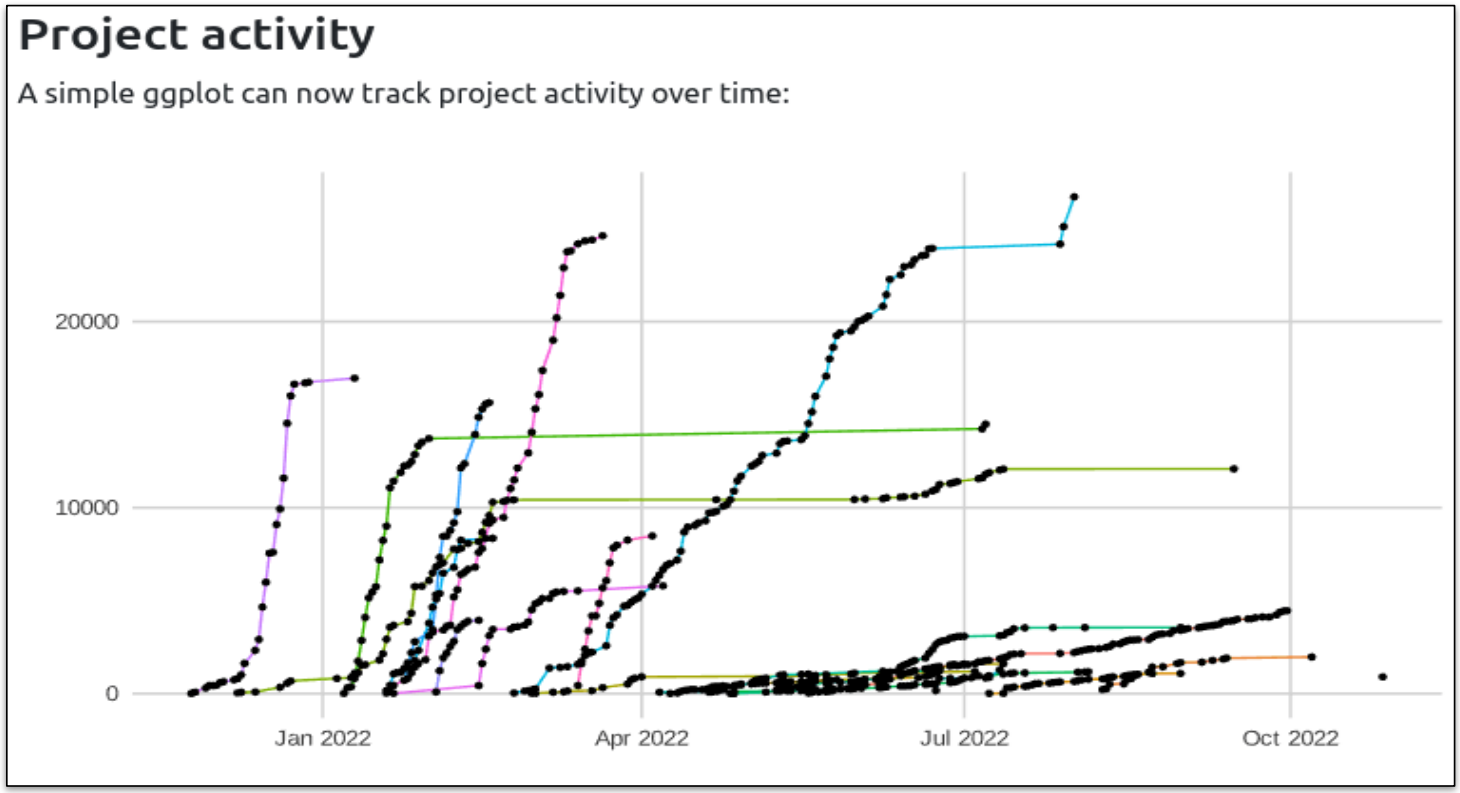
sysrev

Semi-automated systematic review:

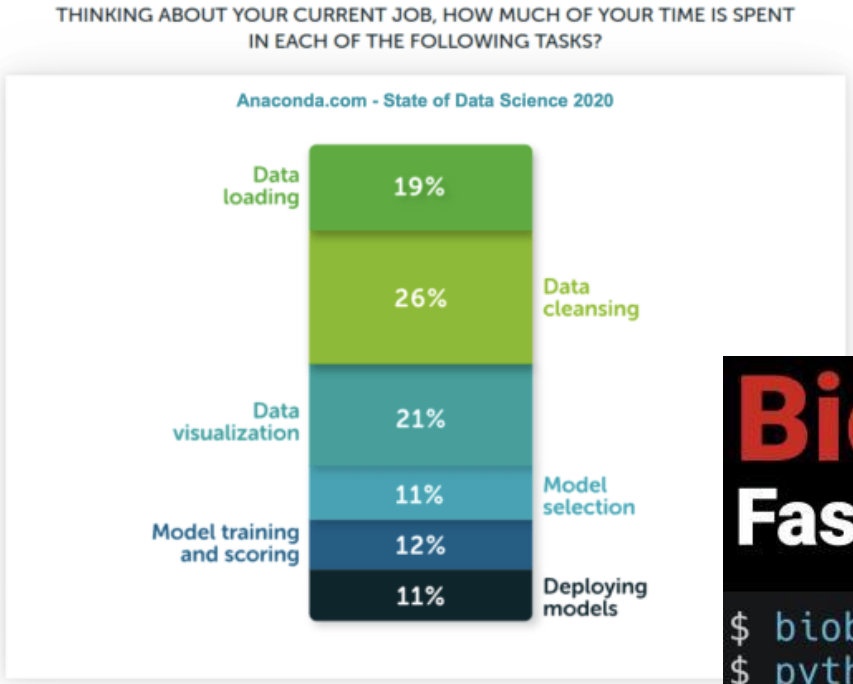
- **Auto-extract /annotate papers**
- **Auto-analyze clustering of papers**
- **Learn from manual inclusion / exclusion**
- **Automated inclusion / exclusion suggestions**
- **Natural Entity Recognition & Causal Relationship Extraction**
- **Feed into ontologies and AI**
- **chatGPT -> bioGPT -> toxGPT (?)**

https://www.youtube.com/channel/UCoUbMAvxBSZpOlqKjOkxNzQ/videos

Data Mining for Ontologies & AI



Progress Dashboard Visualization: Number of *labels* in each ONTOX project over time



45% of time of data analysts is spent loading and cleaning data

BioBricks do this with one-line command

BioBricks.ai
Faster Informatics

```
$ biobricks install tox21
$ python
>>> import biobricks, pandas
>>> tx21 = biobricks.load('tox21')
>>> tx21.tox21.read().load()
#          SAMPLE_ID  PROTOCOL_NAME  ...                SMILES  ...
# NCGC00256074-01  tox21-ache-p3  ...          OCC(=O)OCCCC  ...
# NCGC00255047-01  tox21-ache-p3  ...  Nc1ccc(cc1)C(=O)OCC  ...
# [2075022 rows x 19 columns]
```

Establishment of a big data platform and data gap filling for integration of collected data

- **Biobricks will provide a toolset for hosting, querying, and distributing ONTOX big data.**
- **Biobricks.ai serves sysrev.com data**
- **Biobricks.ai allows brick integrations**
- **~50 BioBricks constructed to date**
- **ChemHarmony:**
  integrates chembl, pubchem, ctdbase etc.:
  200 million triplets of substance/property/result
- **Building querying functionality**
- **Public release of toxicology BioBricks**

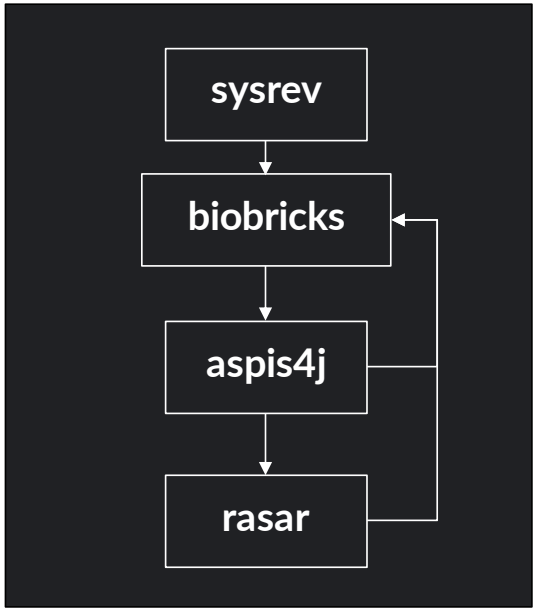
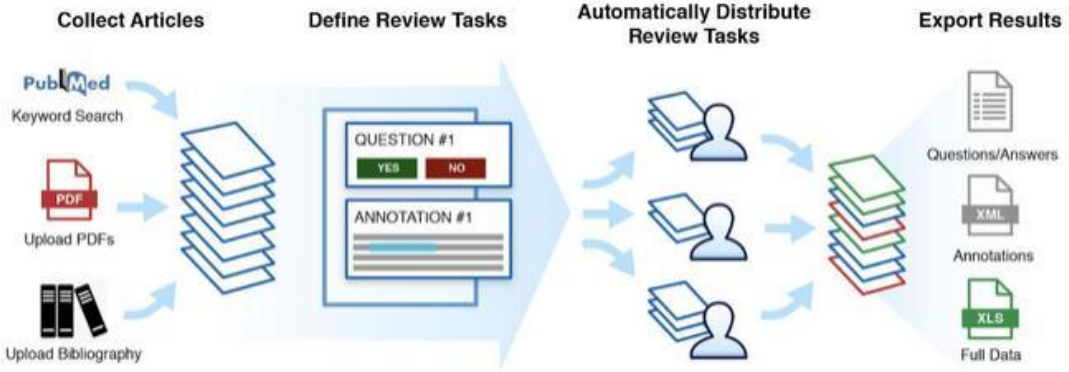


Figure 5.1: Construction and Distribution of ONTOX Data Assets

Literature

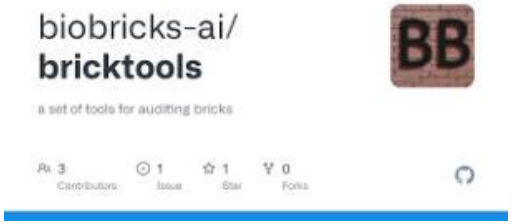
sysrev

https://www.youtube.co
/c/SysRev?app=desktop



Graphic showing the document review workflow on Sysrev.com

Databases



https://www.biopharmatrend.com/post/103-introducing-sysrev-the-intelligent-platform-for-document-review-and-automated-data-extraction/

Internet

Chemchart

http://chemchart.com

DATA

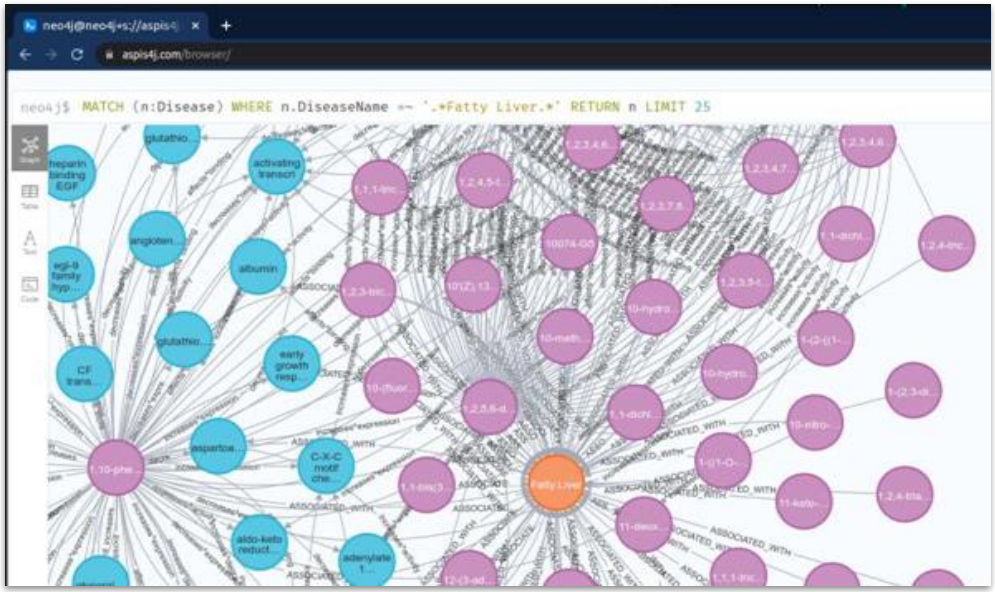
Establishment of a big data platform and data gap filling for integration of collected data

Data integration interoperability

Combine Information obtained by mining literature, public sources, and QSAR predictions

The aspis4j database associates chemicals, proteins, genes, diseases, pathways and other bioinformatics-based entities.

This database will be used to create 'graph embeddings' for predictive, explainable, models.



→ Physiological Maps

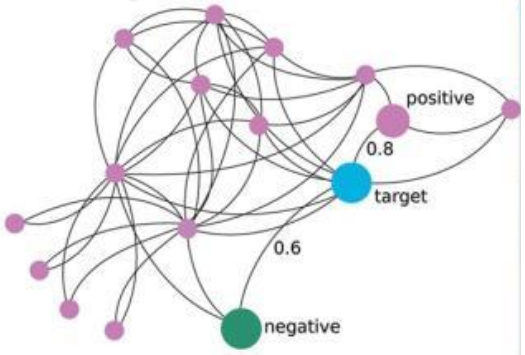
DATA

PMDEP App https://youtu.be/YG0gjm&GD5K

SBML Systems Biology Markup Language

CellDesigner™

SBtab



From perturbation of physiology

Probability of hazard

RASAR + QSAR

From chemical structure and properties

AnImal Replacement

2018: Nine most used animal tests

AI predicted 190,000 chemicals 87% correctly

Animal reproducibility 81%

2020: Human Skin Sensitization

AI predicted 506 chemicals 80% correctly

Animal 74% correct

2022: Nine most used animal tests predicted by AI

AI predicted 4700+ food chemicals 83% correctly in 1h

= 38,000 animal studies at $250+ million

2023: Systemic toxicities

AI predicted 75% cancer risk of 950 chemicals and 82% reproductive tox of 1152 chemicals correctly

Set-up and application of machine learning/deep learning approaches to predict probability of chemical hazard and potency

We can do 1 trillion comparisons per hour on a “normal” computer!

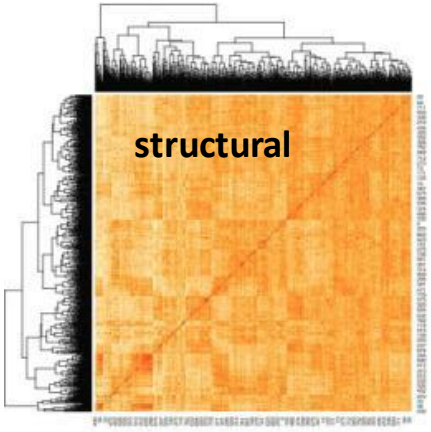


Figure 5.2A. Similarity heatmap for chembert embeddings on 1k chembl compounds
Dark red = highly similar,
White = not similar

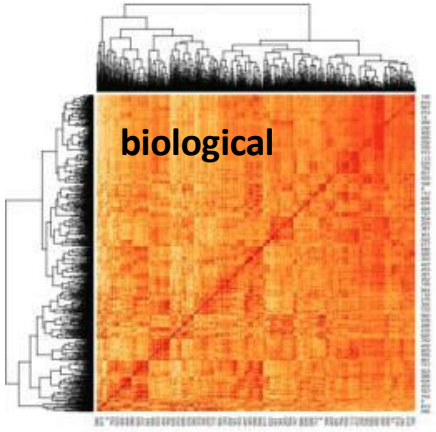
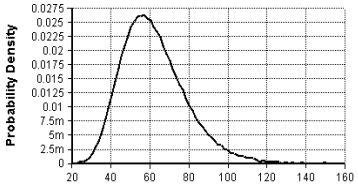
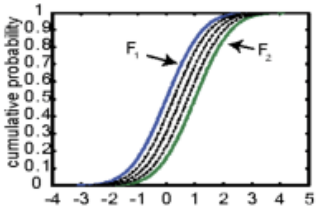


Figure 5.2B. Heatmap for inhibition assay supervised embedding
Dark red = highly similar,
White = not similar

ADME testing



PBPK



Exposure levels (population given chemical)

TTC

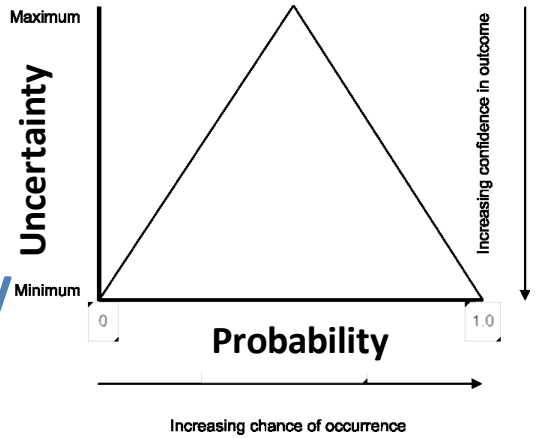
Tissue levels (individuals)

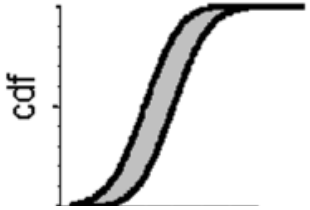
Hartung and Leist, 2008

internal TTC

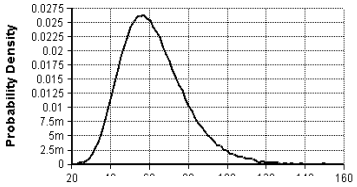
Probability of hazard

Probabilistic Risk (individual or cumulative population)





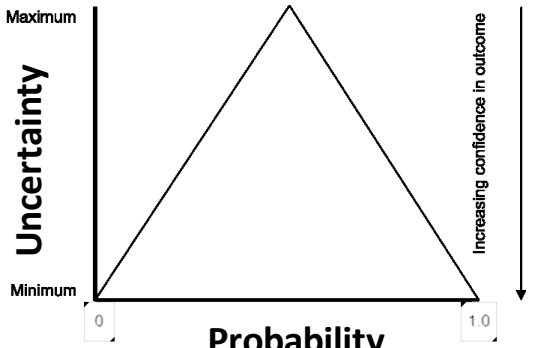
Concentration response



Susceptibility (individuals)

Probability of exposure

Probabilistic Risk (individual or cumulative population)



Minimal risk

More Testing?

Risk Management

Information economics *vs.* precaution

Probability hazard (given chemical)

RASAR

Read-across

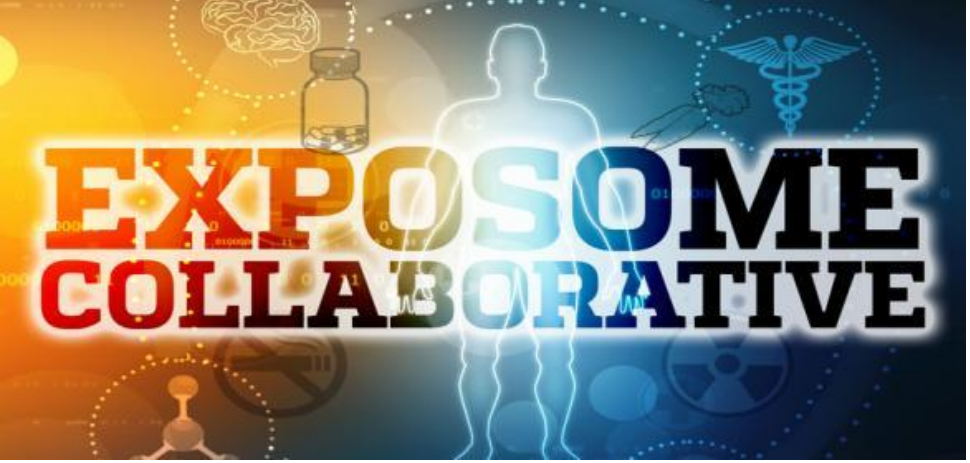
Data

Workshop July 2023

Mechanistic alerts

1. The largest toxicological database ever
2. Automatic paper selection and information extraction
3. Automatic extension of AOPs and Physiological Maps
4. Estimation of internal dose from exposure
5. Probability of hazard from chemical structure and perturbation of biology
6. Probability of risk enabling information economy

Six impossible things
I believe ONTOX will deliver



Fenna Sillé

AI & *exposure*

ALTEX 2020, 37, 3-23

"Progress is impossible without change, and those who cannot change their minds cannot change anything."
George Bernard Shaw (1856-1950)

"If you change the way you look at things, the things you look at change."
Wayne Dyer (1940-2015)

Food for Thought ...

The Exposome – a New Approach for Risk Assessment

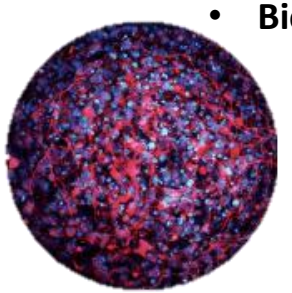
Fenna Sillé[1], Spyros Karakitsios[2], Andre Kleensang[1], Kirsten Koehler[1], Alexandra Maertens[1], Gary W. Miller[3], Carsten Prasse[1], Lesliam Quiros-Alcala[1], Gurumurthy Ramachandran[1], Stephen M. Rappaport[4], Ana M. Rule[1], Denis Sarigiannis[2,5], Lena Smirnova[1] and Thomas Hartung[1,6]

Exposome & AI = EI (Exposome Intelligence)

AI & *in vitro*

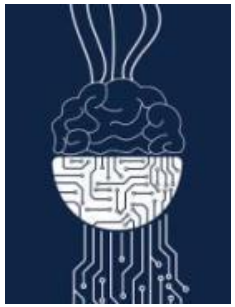
Organoid Intelligence (O.I.)

- Physiology of learning
- Tox & Drug Development
- Biological Computing



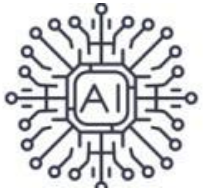
Brain Organoid

O.I.

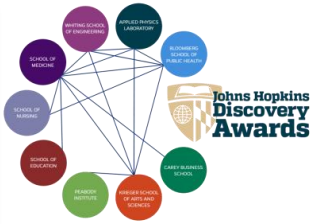


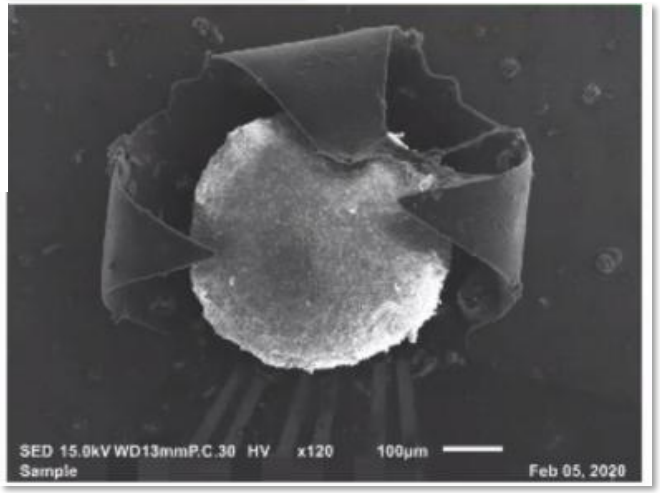
Input

Output



A.I. ⟵





Human brain organoid caged in shell electrodes

frontiers | Frontiers in Science

TYPE Frontiers in Science Lead Article
PUBLISHED 28 February 2023
DOI 10.3389/fsci.2023.1017235

OPEN ACCESS

EDITED BY
Arti Ahluwalia,
University of Pisa, Italy

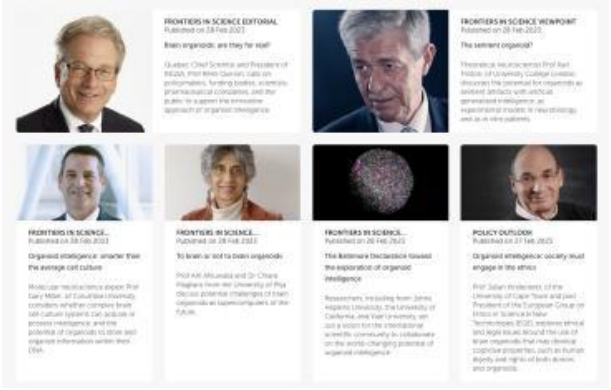
REVIEWED BY
Karl Friston,
University College London,
United Kingdom
Gary Miller,
Columbia University, United States

*CORRESPONDENCE
Thomas Hartung
thartun1@jhu.edu

RECEIVED 11 August 2022
ACCEPTED 07 February 2023
PUBLISHED 28 February 2023

Organoid intelligence (OI): the new frontier in biocomputing and intelligence-in-a-dish

Lena Smirnova[1], Brian S. Caffo[2], David H. Gracias[3,4,5,6,7,8], Qi Huang[3], Itzy E. Morales Pantoja[1], Bohao Tang[2], Donald J. Zack[9], Cynthia A. Berlinicke[10], J. Lomax Boyd[11], Timothy D. Harris[12,13], Erik C. Johnson[14], Brett J. Kagan[15], Jeffrey Kahn[16], Alysson R. Muotri[17,18], Barton L. Paulhamus[19], Jens C. Schwamborn[20], Jesse Plotkin[1], Alexander S. Szalay[21,22,23], Joshua T. Vogelstein[12], Paul F. Worley[24] and Thomas Hartung[1,25]*



frontiers | Frontiers for Young Minds

NEUROSCIENCE AND PSYCHOLOGY
Published: 28 February 2023
doi: 10.3389/frym.2023.1049593

BRAIN-CELL CULTURES: THE FUTURE OF COMPUTERS AND MORE?

Lena Smirnova, Itzy Erin Morales Pantoja and Thomas Hartung*

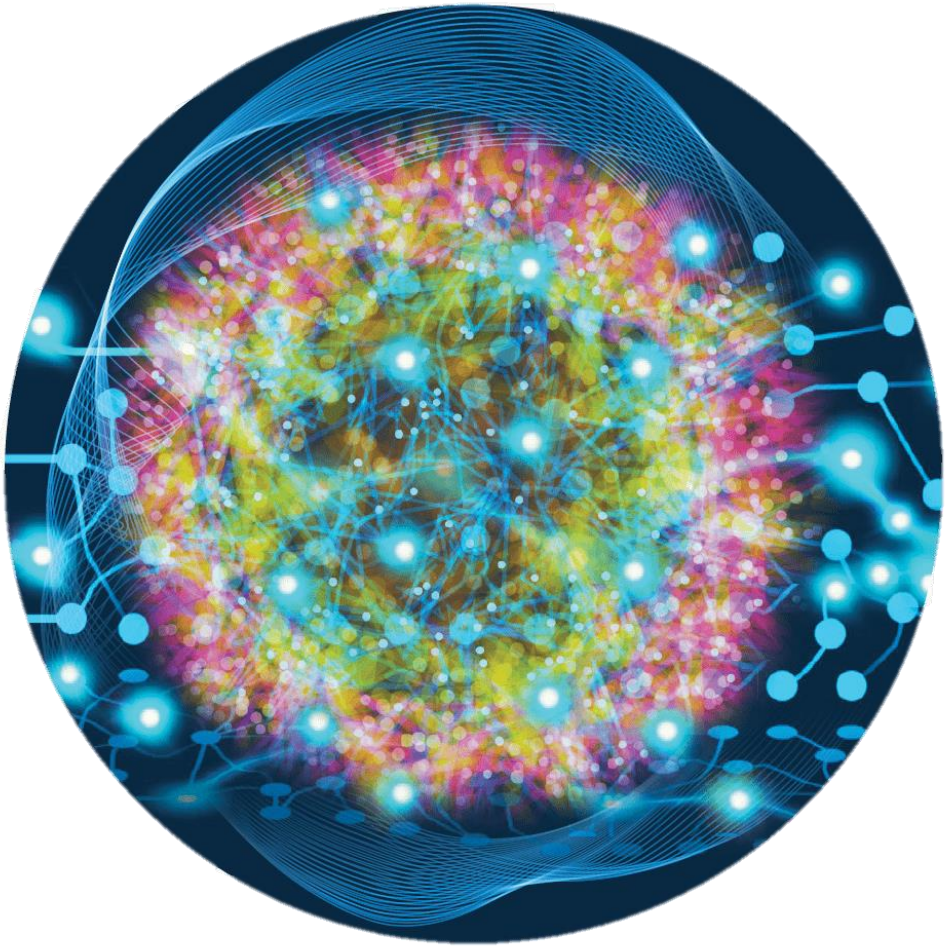
Lay summary

Front.Sci., 27 Feb 2023
10.3389/fsci.2023.1017235
This is part of an article hub

Explore article hub →

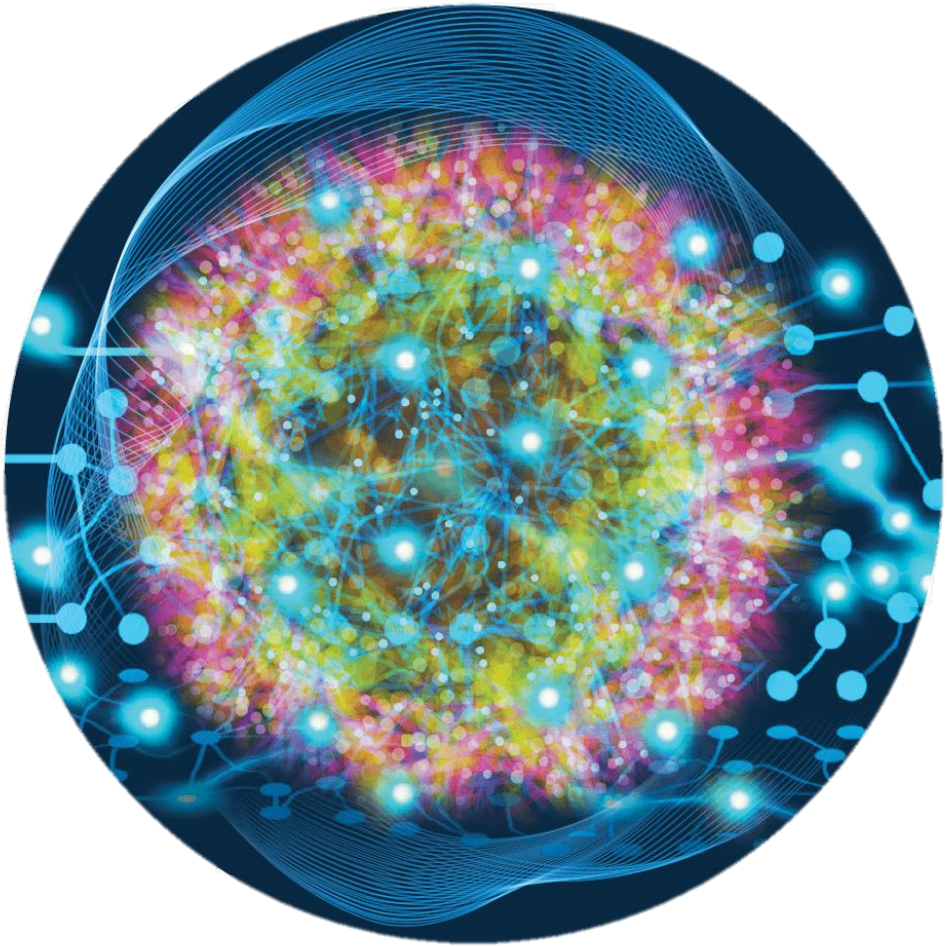
Read lead article →

Powering up the next generation of biocomputers with brain organoids



ToxAIcology

ALTERNATIVES TO ANIMAL TESTING



ToxAIcology

ALTERNAT**I**VES

T**O**AN**I**MAL

T**E**ST**I**NG