

From Connectomes to Digital Twins: Forecasting the Brain in Real Time

Mapping the Living Mind: From Wiring Diagrams to Neural Forecasting

Scientists have spent years trying to figure out how the biological brain works by looking at it from two different angles. One group has focused on connectomics, which is basically mapping the physical wiring of the brain. The other group has looked at functional imaging, or watching neurons fire in real time. We are now seeing these two fields merge through advanced AI to create what researchers call a digital twin of the brain. This move goes beyond just taking high-resolution pictures. It is about building models that can actually predict what a brain will do next.

Building the Physical Maps

The foundation of this work is the wiring diagram. We recently saw a massive milestone with the completion of the central brain connectome for the adult fruit fly, *Drosophila melanogaster*. This map includes more than 125,000 neurons and 50 million synaptic connections. While a fly brain is small, the data is incredibly complex. A single neuron might connect to hundreds of others, making it very difficult to understand how these paths lead to specific behaviors.

We are seeing similar progress in humans too. Researchers recently reconstructed a tiny fragment of the human cerebral cortex. Even though it was only one cubic millimeter in size, it required over a petabyte of data to map at a nanoscale resolution. These physical maps have shown us things we never knew existed, like neurons that form unusual triangular shapes. However, as many experts have pointed out, a connectome is just a map. It does not tell us how the “traffic” of neural activity moves through those wires.

Predicting the Traffic of the Brain

To solve this, researchers are turning to neural forecasting. One of the most important tools in this area is the Zebrafish Activity Prediction Benchmark, or ZAPBench. It uses light sheet microscopy to record the activity of over 70,000 neurons in larval zebrafish. This is currently the only vertebrate where we can see the whole brain active at once at such a high resolution.

By using models originally built for weather forecasting, like those in WeatherBench, scientists are testing how well AI can predict the next 30 seconds of a brain’s activity based on just a few seconds of history. This is a massive shift in how we study neuroscience. Instead of just describing what happened, we are trying to forecast what will happen.

Several new techniques are making this possible:

- **Volumetric Video Models:** Instead of just looking at individual neuron signals, new models like 4D UNets

look at the raw 3D video over time. This helps the AI understand the spatial relationships between neurons that other methods might miss.

- **Foundation Models:** Just like the models that power modern chat tools, new foundation models of the mouse visual cortex are being trained on huge amounts of data. These models can be applied to new animals they have never seen before, successfully predicting how their neurons will react to new videos.
- **Classification Strategies:** New architectures like QuantFormer are changing the way we think about brain signals. Instead of trying to predict a continuous wave of activity, they treat neural spikes like a classification problem. This has proven much more effective at capturing the quick, sparse bursts of energy that define how neurons communicate.

Why Global Brain States Matter

One of the biggest hurdles in this research is that a single neuron does not act alone. Its behavior is often influenced by the global state of the brain, such as whether an animal is alert or performing a specific task. A model called POCO, which stands for Population Conditioned forecaster, handles this by looking at local neuron dynamics while also considering the overall state of the entire population. This helps the model understand how shared brain structures influence individual cells.

Future Applications and Interventions

The goal of this research is not just to understand the brain but to interact with it. If we can forecast neural activity in real time, we can develop systems that intervene before something goes wrong. Some models can now run in as little as 3.5 milliseconds. This speed could allow for closed-loop optogenetic interventions, where light is used to stimulate neurons to stop a seizure or a specific craving before the person even realizes it is happening.

We are moving into an era where we can see inside ourselves with the same clarity that we see the world around us. While managing petabytes of data is a major challenge, combining physical maps with AI forecasting brings us much closer to a true mechanistic understanding of intelligence.

This post was written with the help of AI for analysis, using the NotebookLM shared resource here:

<https://notebooklm.google.com/notebook/74dc7f14-54cb-481b-9ee8-8347a6f5cba1>

References and Research Links

- **A Drosophila computational brain model reveals sensorimotor processing**
<https://doi.org/10.1038/s41586-024-07763-9>

- **A connectome and analysis of the adult *Drosophila* central brain** <https://doi.org/10.7554/eLife.57443>
 - **A petavoxel fragment of human cerebral cortex reconstructed at nanoscale resolution** <https://doi.org/10.1126/science.adk4858>
 - **Foundation model of neural activity predicts response to new stimulus types** <https://doi.org/10.1038/s41586-025-08829-y>
 - **POCO: Scalable Neural Forecasting through Population Conditioning** <https://arxiv.org/abs/2410.18025>
 - **QuantFormer: Learning to Quantize for Neural Activity Forecasting** <https://arxiv.org/abs/2405.17140>
 - **ZAPBench: A Benchmark for Whole-Brain Activity Prediction in Zebrafish** <https://openreview.net/forum?id=oCHsDpyawq>
 - **Forecasting Whole-Brain Neuronal Activity from Volumetric Video** <https://arxiv.org/abs/2503.00073>
 - **A connectome is not enough - what is still needed to understand the brain of *Drosophila*?** <https://doi.org/10.1242/jeb.242740>
-

Digital Twin - exploring the basics

The concept of digital twins is not new, but rather built on ideas that have been explored for the last couple of decades. The technology (compute power, data management & analytics, etc..) and thinking (increasing regulatory and community acceptance of digital approaches to science) have finally hit an inflection point that makes *in silico* modeling attainable in a cost effective manner.

What this now unlocks is a new opportunity set in the form of machine accessible data, as well as integration of the data sets / ontologies across the target systems / interactions. The need to get to a standardized mechanism to make these data available is tied to the FAIR Data work, and an important dimension to Digital Twin.

Digital twins vs. simulations

Although simulations and digital twins both utilize digital models to replicate a system's various processes, a digital twin is actually a virtual environment, which makes it considerably richer for study. The difference between digital twin and simulation is largely a matter of scale: While a simulation typically studies one particular process, a digital twin can itself run any number of useful simulations in order to study multiple processes.

SOURCE: [IBM](#) , [WHAT IS A DIGITAL TWIN](#)

At its heart, the idea of a digital twin is to reproduce a system in a “runnable” computer model. This oversimplifies the idea, but is a useful construct to think about the problem space and the opportunity it presents. If you can take a scientific instrument, and fully model it in silico, you can then run data sets through it virtually – this makes the assumption that both the inbound and outbound data are available in a machine usable format – something that is tied to this work.

Digital twin is an interdisciplinary research field which includes engineering, computer science, automation and control, and so on. But due to the multidisciplinary nature of the field, it also touches on materials science, communication, operations management, robotics, medicine and other disciplines. A keyword analysis indicates that digital twin, ‘smart manufacturing’, ‘big data’, ‘cyber-physical system’, and ‘digital economy’ are closely related fields.

SOURCE: “INNOVATIONS IN DIGITAL TWIN RESEARCH” FROM [NATURE PORTFOLIO](#)

The article in nature.com is an interesting piece in that it ties together the many dimensions in this field of research. We can’t think of “Digital Twin” as a single entity opportunity, rather to fully realize the potential, we need to look at it as a part of an emerging “virtual capability ecosystem” with applications back to the real world. The value is realized in lower long term costs with increased innovation driven by reduced cost and cycle times, accompanied by increases in application of AI / ML on these models to gain targeted insights that more sharply focus the bench work.

Track the past and help predict the future of any connected environment

SOURCE: AZURE DIGITAL TWINS

The ability to create learning models for these Digital Twins will improve the accuracy and usefulness of the models over time, and that feedback loop will be a critical part of design. While the industry is maturing, we are seeing more vendors coming to the table with solutions in this space. One of the interesting things to watch is how we as an industry continue to drive open standards in support of these ideas to avoid the traps of “vendor lock in” that were so prevalent in the past.

Pistoia Alliance: Patient Centricity

There is an increasing recognition of the value in patient engagement with respect to healthcare in general, as well as the emerging field of personalized / targeted medicine and digital health. The wearable / therapeutic combination, CAR-T therapies, telehealth and so much more fall into this broad category of patient centricity

and experience, as well as the direct marketing side of it.

The
Pisto
oia
Allia
nce
has
call
ed
for
life
scie
nce
and
heal
thca
re
to
urg
entl
y
rest
ruct
ure
aro
und
pati
ent
cent
ricit
y -
rea
d
the
post
fro
m
the
allia
nce
her
e.



the pandemic has changed behaviors. Billions of people changed the way they interact with healthcare in a matter of months. In this new era of targeted precision medicine, we all play a role in creating the patient-centric future that patients deserve.”

**CRISTINA ORTEGA DURAN, CHIEF DIGITAL HEALTH OFFICER R&D
FOR ASTRAZENECA**

I am excited to see where this leads us as an industry, and how we shift from traditional approaches to include our broad patient populations in developing and delivering medicines and treatments. It will be great to see growing inclusivity across geographic and social boundaries as we increase reach and engagement.

Is your Scientific Data FAIR

For many years, we have seen the proliferation of data as we increasingly instrument our scientific processes. We have developed a diverse landscape of tools and processes, making significant leaps from paper based documentation, but created a new nightmare of integration and complex analysis. The FAIR initiative or set of principles is a framework to reduce that complexity through the application of a core set of principles outlined below, making data machine readable across sources. This unlocks the data from the proprietary structure and system walls, and offers a foundation to build interconnected analysis and insights.

Reference this excerpt from the abstract here that summarizes quite nicely what the objective is:

There is an urgent need to improve the infrastructure supporting the reuse of scholarly data. A diverse set of stakeholders—representing academia, industry, funding agencies, and scholarly publishers—have come together to design and jointly endorse a concise and measurable set of principles that we refer to as the FAIR Data Principles. The intent is that these may act as a guideline for those wishing to enhance the reusability of their data holdings. Distinct from peer initiatives that focus on the human scholar, the FAIR Principles put specific emphasis on enhancing the ability of machines to automatically find and use the data, in addition to supporting its reuse by individuals.

[HTTPS://WWW.NATURE.COM/ARTICLES/SDATA201618#ABS1](https://www.nature.com/articles/sdata201618#ABS1)

The FAIR Guiding Principles

To be Findable:

- F1. (meta)data are assigned a globally unique and persistent identifier
- F2. data are described with rich metadata (defined by R1 below)
- F3. metadata clearly and explicitly include the identifier of the data it describes
- F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:

- A1. (meta)data are retrievable by their identifier using a standardized communications protocol
 - A1.1 the protocol is open, free, and universally implementable
 - A1.2 the protocol allows for an authentication and authorization procedure, where necessary
- A2. metadata are accessible, even when the data are no longer available

To be Interoperable:

- I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (meta)data use vocabularies that follow FAIR principles
- I3. (meta)data include qualified references to other (meta)data

To be Reusable:

- R1. meta(data) are richly described with a plurality of accurate and relevant attributes
 - R1.1. (meta)data are released with a clear and accessible data usage license
 - R1.2. (meta)data are associated with detailed provenance
 - R1.3. (meta)data meet domain-relevant community standards

There are an increasing number of resources targeted at supporting the movement to FAIR data, a couple of which are included here to get you started. There is much to cover on this topic, but these links and materials are a start on the conversation.

[How to GO FAIR](#)

<https://www.pistoiaalliance.org/projects/current-projects/fair-implementation/>

References

- NIST:
<https://www.nist.gov/itl/ssd/information-systems-group/configurable-data-curation-system-cdcs/cdcs-help-an>

d-resources-1

- scientific data (Nature.com): <https://www.nature.com/articles/sdata201618#Abs1>
 - Local link in case above link fails: [The FAIR Guiding Principles for scientific data management and stewardship](#)
-

GDPR compliance steps, and impacts on MDM thinking

The impacts of the General Data Privacy Regulations are tremendous and cannot be overstated for any business that deals in personal or private data. In large enterprises, the subject of Master Data Management is often a 3rd rail, or considered a sure path to career death due to the complexity of the topic. Done properly the benefits are tremendous, but the ownership of information at the functional level is so often a barrier to success that it has become generally understood that these efforts are rarely fully successful.

Now we enter the world of GDPR. The regulation in principle is simple. It can be generalised to say: If you collect any personal information (broadly defined) you must make the information owner aware of what you collect, what you do with it and how to get that information back or purge it.

While it sounds simple in application, the complexity of the average enterprise information ecosystem makes this horrendously challenging. In the pharmaceutical industry where I currently work, patient information is collected in the course of a number of activities. As a part of that, various systems and reports are impacted and the process gets exponentially more complicated when biospecimens (samples of tissue, blood, etc...) are concerned. All samples must have consent for use documented, and a traceability through the full lifecycle. When data is “blinded” or anonymized, this adds even more complexity as the ramifications of “unblinding” data are significant to the integrity of the overall trial process. Beyond patient data, there is data on clinicians, research organizations, collaborators and more.

My recommendations I have made, and I suggest to clients include the following:

1. A comprehensive audit of all Personally Identifiable Information collected or generated in the course of business. This should result in a list of data attributes, source of collection, systems and processes impacted
2. For all data collected, where mastering or simplification of process is feasible, a plan must be defined and should be mapped into an overall value stream analysis for inclusion in a portfolio of work focused on GDPR compliance.
3. All data once mapped, must be included in a reference that allows comprehensive aggregation of the record - unifying the person to a single view. This requirement is where the value of mastering comes to the top. Without the ability to view a person in a record, the next step is exponentially more challenging.
4. The Data Privacy Officer for the company and if applicable, related functions, must be pulled in to evaluate

the plans and approach, supplementing the data record or approach as needed.

5. The next step in compliance once the data is clearly mapped and understood is an evaluation of the impact of “providing and purging” as each data owner has the right to request at any time a comprehensive export of all data collected about them. Correspondingly, they can also request that all personal data be purged. This becomes particularly challenging when data has cross dependencies and is a part of a larger analysis.
6. The evaluation of the provide and purge process must result in a documented program and process to answer the question of what is the impact of a request. How will we mitigate the risk and what needs to change about our processes to accommodate these needs. This starts with the evaluation, moves to a documented plan, and then must be accompanied by at a minimum, a tabletop exercise of the multiple scenarios identified.

This is a costly effort, from a time and resources perspective. As an enterprise, leaders must be aware of the risk of non compliance relative to the cost of compliance.

Mapping antibiotic resistance with the CDDEP

The Center for Disease Dynamics, Economics & Policy (CDDEP) provides an interactive map here:<http://resistancemap.cddep.org/>

The data are illuminating and should cause some concern as we look at the growing resistance. There is also hope in noting the prevention of this issue to a large degree by key countries where policy is enforced and use is seen differently, with the consumer not driving the “drug bus”, rather an informed physician guided by long range policy.

Hotbed Of Antibiotic Resistance

We have to start globally funding increased R&D here before it's too late. Over time, due to heavy use and natural evolution of bacteria, we have seen a significant increase in resistant strains that defy standard treatment courses. This is coupled with regulatory approval pressures and economic pressures that are reducing the innovation and funding for new antibiotic research. As these two trends converge, a crisis is brewing that we need to address.

Read a related article here: <http://n.pr/1KmJoth>