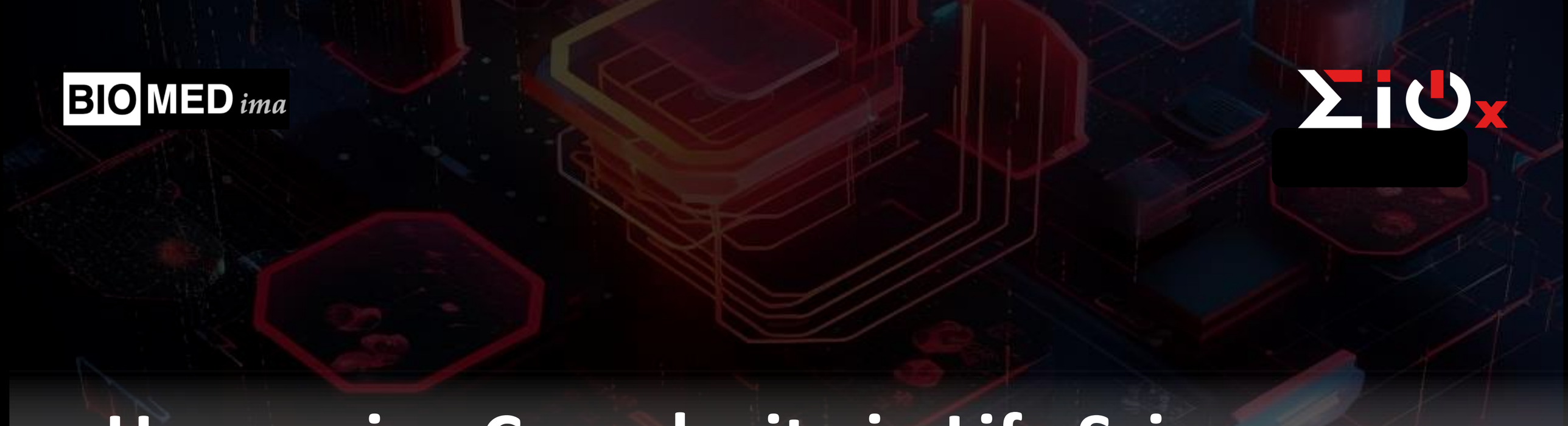
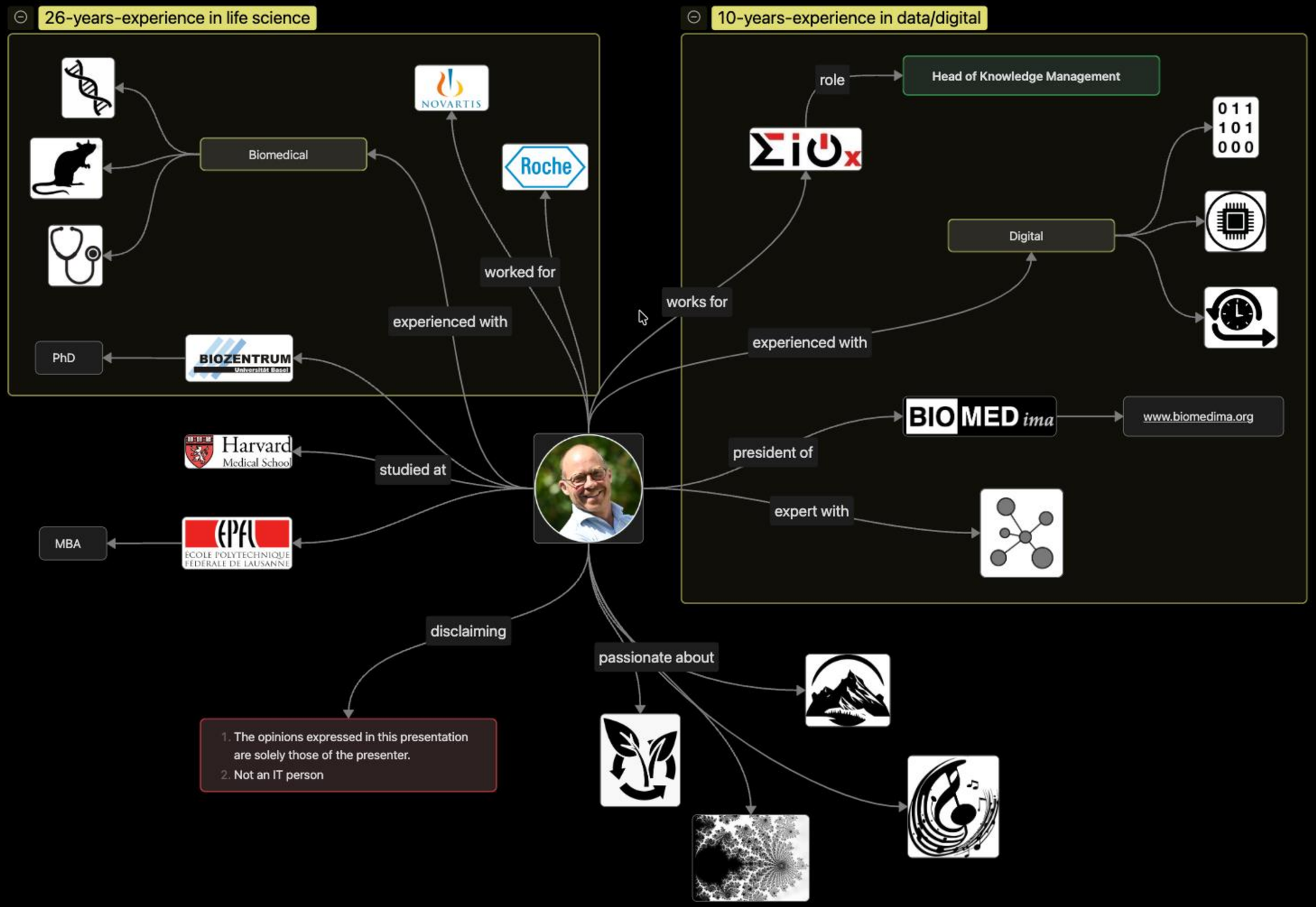


Harnessing Complexity in Life Sciences

A Problem-Solving Framework aiming at Digital Sobriety and Sovereignty

Cedric Berger, PhD, MBA

2026-03-25


26-years-experience in life science
Biomedical
NOVARTIS
Roche
worked for
experienced with
PhD
BIOZENTRUM
Universität Basel
Harvard
Medical School
studied at
MBA
EPFL
ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE
10-years-experience in data/digital
role
Head of Knowledge Management
works for
Digital
experienced with
BIOMEDima
www.biomedima.org
president of
expert with
disclaiming
1. The opinions expressed in this presentation are solely those of the presenter.
2. Not an IT person
passionate about

Before We Begin

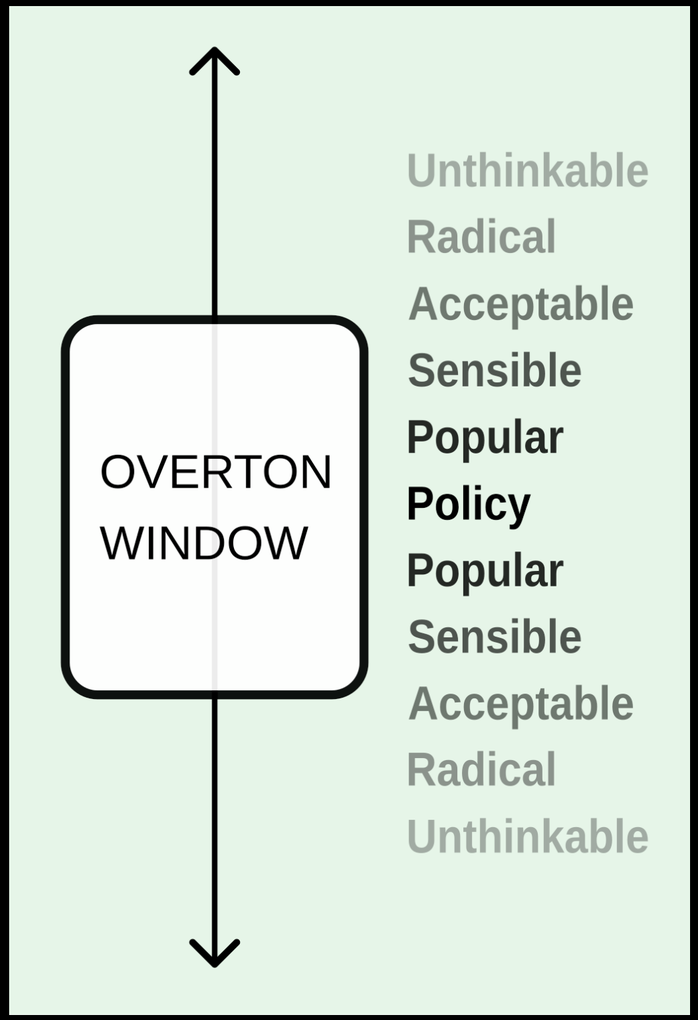
In this presentation you may,

- hear ideas that challenge your current beliefs
- encounter perspectives that conflict with your experience
- feel moments of discomfort, disagreement, or surprise

I hope you agree with the following:

- curiosity shall prime over judgment
- discomfort is not danger
- you don’t have to agree
- growth happens outside the comfort zone

> Take a breath. Let’s explore what’s possible, together.



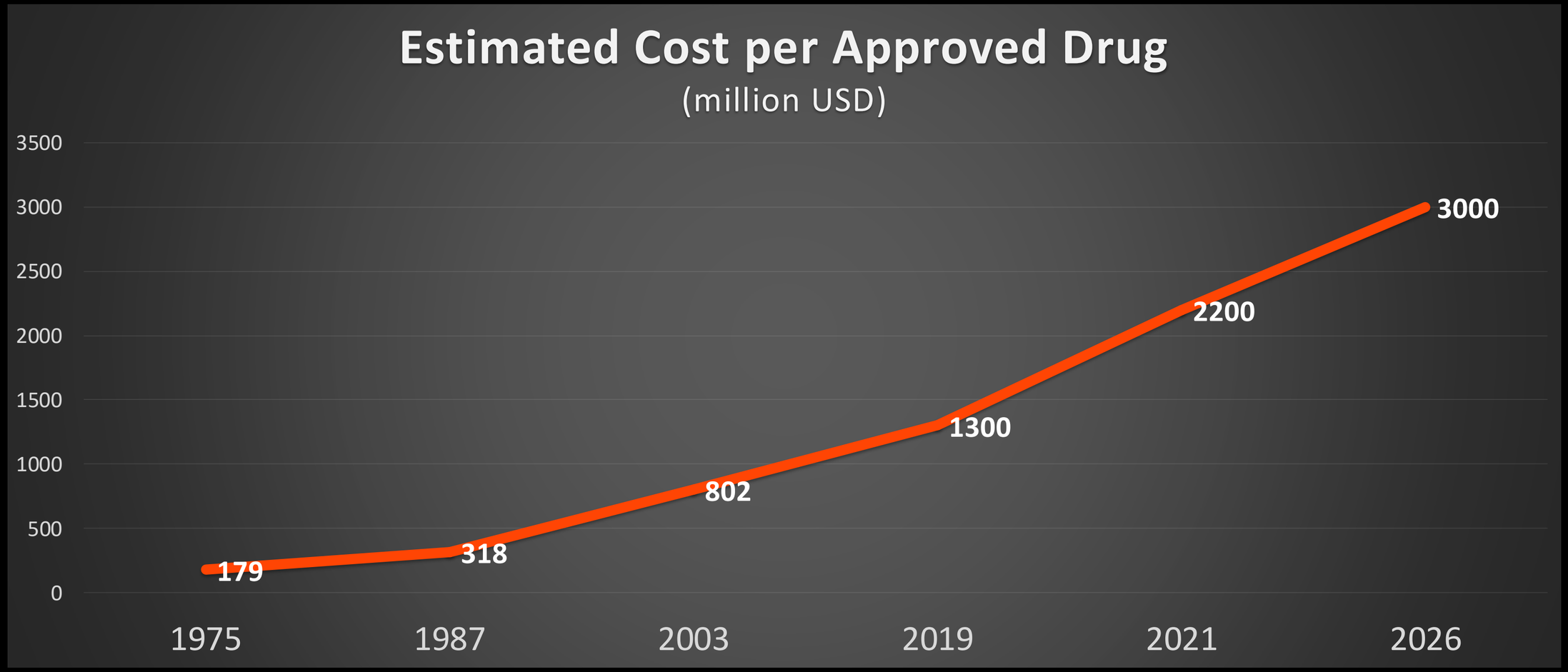
https://en.wikipedia.org/wiki/Overton_window

Context

Life Science (and beyond)

1. Digital Complexity

Cost of Drug Development



Blended from multiple sources:
Deloitte: https://www.genengnews.com/gen-edge/the-unbearable-cost-of-drug-development-deloitte-report-shows-15-jump-in-rd-to-2-3-billion/
JAMA network: https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2828689
TUFTS CSDD: https://greenfieldchemical.com/2023/08/10/the-staggering-cost-of-drug-development-a-look-at-the-numbers/
Springer: https://link.springer.com/article/10.1007/s43441-024-00667-w

Raising Inefficiencies

The current, inherited, paper-based and milestone-driven RDM&C[1] process is highly inefficient.

- taking 10 to 15 years
- costing $2.6 billion

to develop and market a single medicine[2], especially during the development phase:

- clinical trials have a 7.9% success rate[3]

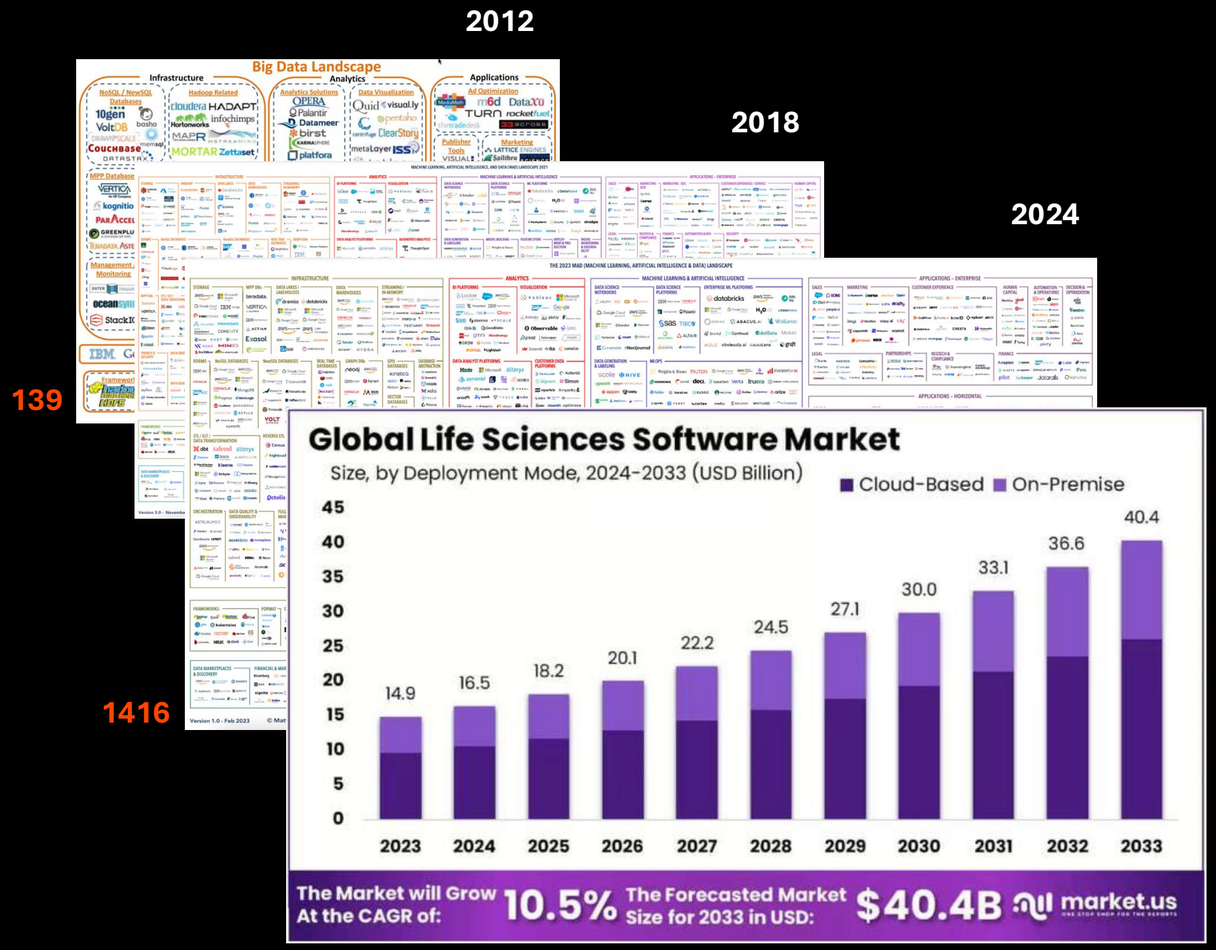
The pharmaceutical industry struggles to reinvent itself due to

- inherent business complexities
- increasing regulations
- a complaisant and risk-averse culture
- reluctancies to adopt digital practices

1) Research, Development, Manufacturing and Commercialization; 2) and 3) Tufts CSDD benchmark

Mostly Due to Increase of Digital Complexity

Mostly Due to the Proliferation of IT Systems



https://mattturck.com

https://market.us/report/life-sciences-software-market/

Depending on definitions of “big” and “IT system”, big pharma organizations *

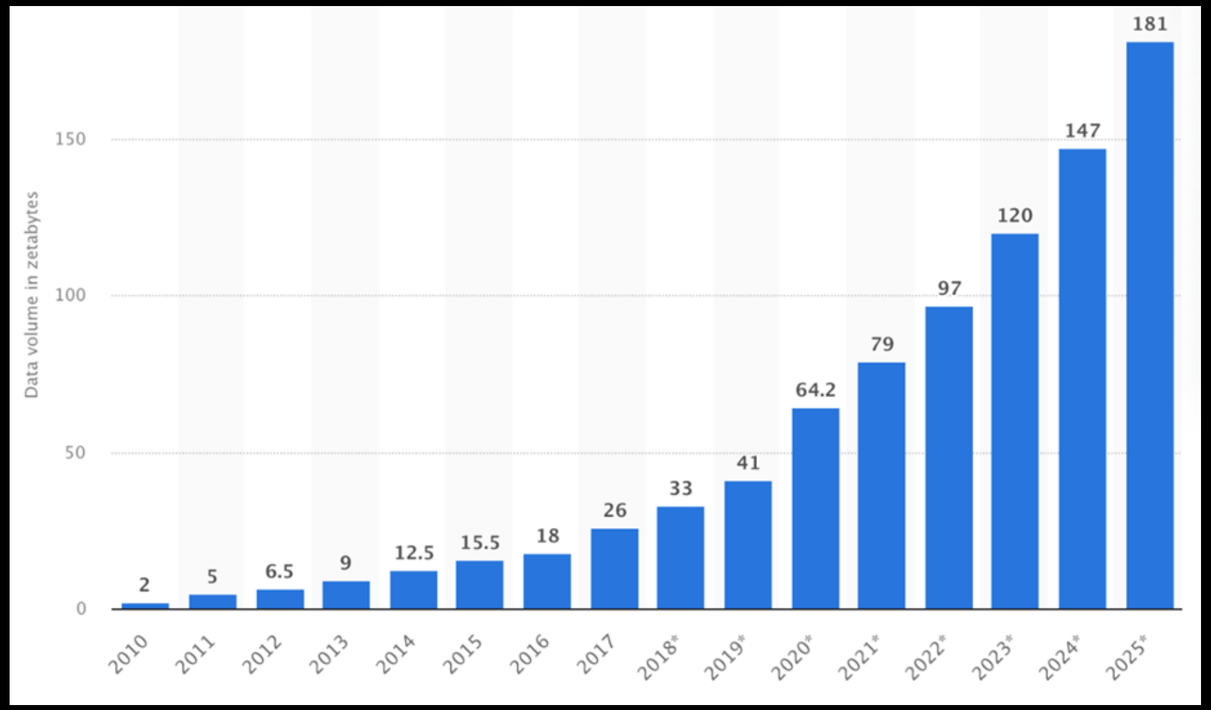
- operate anywhere from **500 to over 6’000 IT systems**
- handling **100 petabytes to several exabytes annually**

100 petabytes = 25 million DVDs
1 exabyte = 1000 petabytes = 250 million DVDs

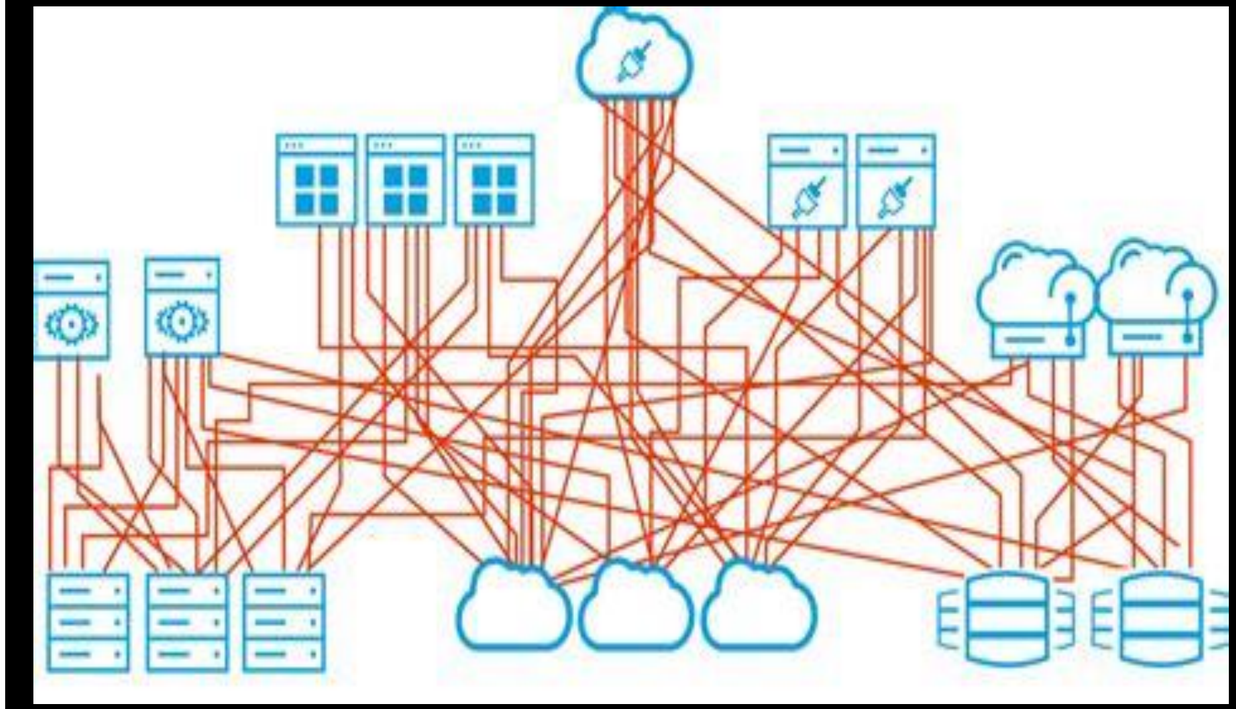
* Estimated from multiple sources: https://www.pharma-iq.com/manufacturing/articles/navigating-ai-integration-in-pharmas-legacy-systems; https://www2.deloitte.com/us/en/insights/industry/life-sciences/biopharma-digital-transformation.html; https://moldstud.com/articles/p-exploring-systems-engineering-practices-in-the-pharmaceutical-industry; https://www.pharmalex.com/thought-leadership/blogs/key-considerations-for-data-strategies-in-the-pharmaceutical-industry/

Leading to Big, Siloed, Disintegrated, Contextless data

Volume of data/information (zetabytes) , created, captured, copied, consumed worldwide from 2010 to 2025.



Spaghetti-monster data integration from point-to-point systems



Statista: https://www.statista.com/

Data Quality has Never Been a Priority

Traditional Approaches don’t Work Anymore

Business Expert

“Our patients' data spans over 3000 tables”

“EHR systems give me incomplete, outdate, discrepant data about the same patients?”

Knowledge Expert

“Then, you will never know what is going on with your patients”

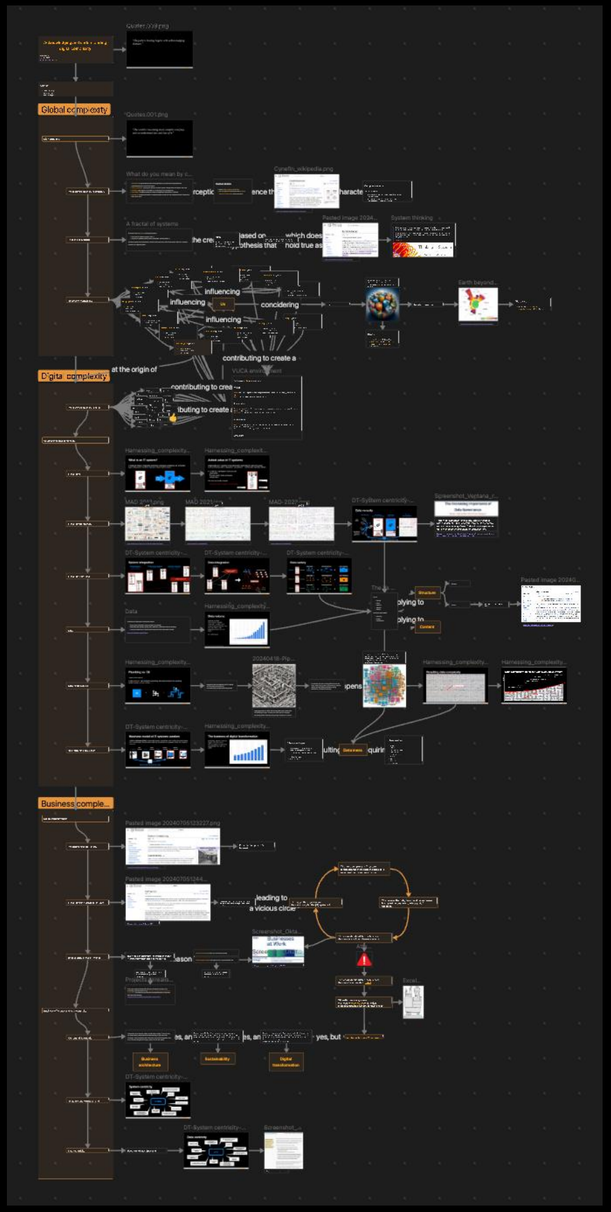
“These systems haven’t been designed to communicate together and exchange data.”

“It is likely that nobody understands the business end-to-end anymore”

More about Digital Complexity

BIOMED ima

https://www.biomedima.org



https://www.biomedima.org/project/acknowledging-and-understanding-digital-complexity/

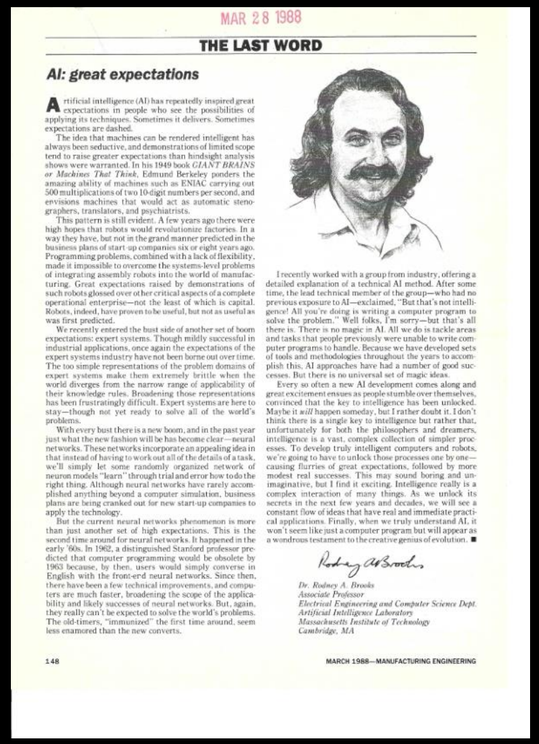
Context

Life Science (and beyond)

2. Artificial Intelligence

Not Everyone will Benefit from AI

The Hype



MAR 28 1988

THE LAST WORD

AI: great expectations

Artificial intelligence (AI) has repeatedly inspired great expectations in people who see the possibilities of applying its techniques. Sometimes it delivers. Sometimes expectations are dashed.

The idea that machines can be rendered intelligent has always been seductive, and demonstrations of limited scope tend to raise greater expectations than hindsight analysis shows were warranted. In his 1949 book *GIANT BRAINS or Machines That Think*, Edmund Berkeley ponders the amazing ability of machines such as ENIAC carrying out 500 multiplications of two 10-digit numbers per second, and envisions machines that would act as automatic stenographers, translators, and psychiatrists.

This pattern is still evident. A few years ago there were high hopes that robots would revolutionize factories. In a way they have, but not in the grand manner predicted in the business plans of start-up companies six or eight years ago. Programming problems, combined with a lack of flexibility, made it impossible to overcome the systems-level problems of integrating assembly robots into the world of manufacturing. Great expectations raised by demonstrations of such robots glossed over other critical aspects of a complete operational enterprise—not the least of which is capital. Robots, indeed, have proven to be useful, but not as useful as was first predicted.

We recently entered the bust side of another set of boom expectations: expert systems. Though mildly successful in industrial applications, once again the expectations of the expert systems industry have not been borne out over time. The too simple representations of the problem domains of expert systems make them extremely brittle when the world diverges from the narrow range of applicability of their knowledge rules. Broadening those representations has been frustratingly difficult. Expert systems are here to stay—though not yet ready to solve all of the world's problems.

With every bust there is a new boom, and in the past year just what the new fashion will be has become clear—neural networks. These networks incorporate an appealing idea in that instead of having to work out all of the details of a task, we'll simply let some randomly organized network of neuron models "learn" through trial and error how to do the right thing. Although neural networks have rarely accomplished anything beyond a computer simulation, business plans are being cranked out for new start-up companies to apply the technology.

But the current neural networks phenomenon is more than just another set of high expectations. This is the second time around for neural networks. It happened in the early '60s. In 1962, a distinguished Stanford professor predicted that computer programming would be obsolete by 1963 because, by then, users would simply converse in English with the front-end neural networks. Since then, there have been a few technical improvements, and computers are much faster, broadening the scope of the applicability and likely successes of neural networks. But, again, they really can't be expected to solve the world's problems. The old-timers, "immunized" the first time around, seem less enamored than the new converts.

I recently worked with a group from industry, offering a detailed explanation of a technical AI method. After some time, the lead technical member of the group—who had no previous exposure to AI—exclaimed, "But that's not intelligence! All you're doing is writing a computer program to solve the problem." Well folks, I'm sorry—but that's all there is. There is no magic in AI. All we do is tackle areas and tasks that people previously were unable to write computer programs to handle. Because we have developed sets of tools and methodologies throughout the years to accomplish this, AI approaches have had a number of good successes. But there is no universal set of magic ideas.

Every so often a new AI development comes along and great excitement ensues as people stumble over themselves, convinced that the key to intelligence has been unlocked. Maybe it *will* happen someday, but I rather doubt it. I don't think there is a single key to intelligence but rather that, unfortunately for both the philosophers and dreamers, intelligence is a vast, complex collection of simpler processes. To develop truly intelligent computers and robots, we're going to have to unlock those processes one by one—causing flurries of great expectations, followed by more modest real successes. This may sound boring and unimaginative, but I find it exciting. Intelligence really is a complex interaction of many things. As we unlock its secrets in the next few years and decades, we will see a constant flow of ideas that have real and immediate practical applications. Finally, when we truly understand AI, it won't seem like just a computer program but will appear as a wondrous testament to the creative genius of evolution. ■

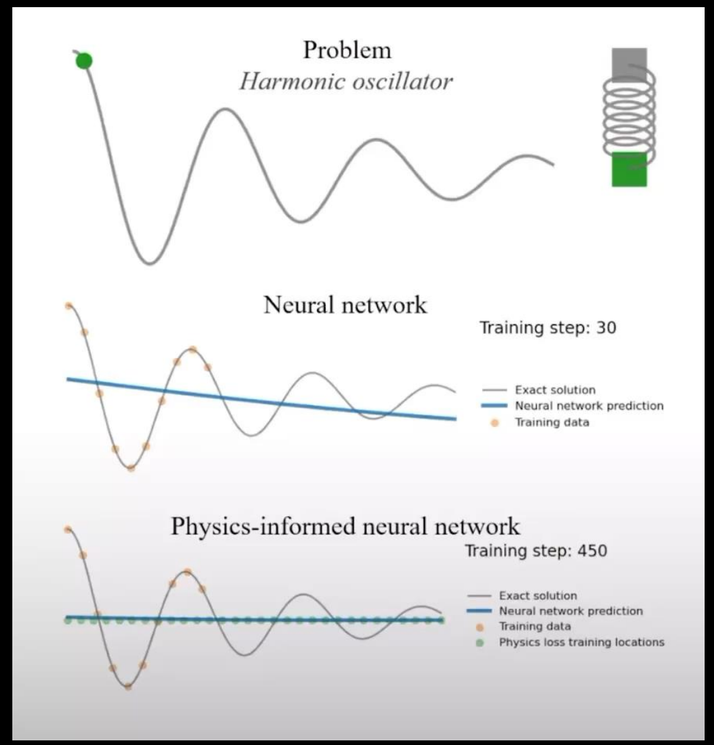
Rodney A. Brooks

Dr. Rodney A. Brooks
Associate Professor
Electrical Engineering and Computer Science Dept.
Artificial Intelligence Laboratory
Massachusetts Institute of Technology
Cambridge, MA

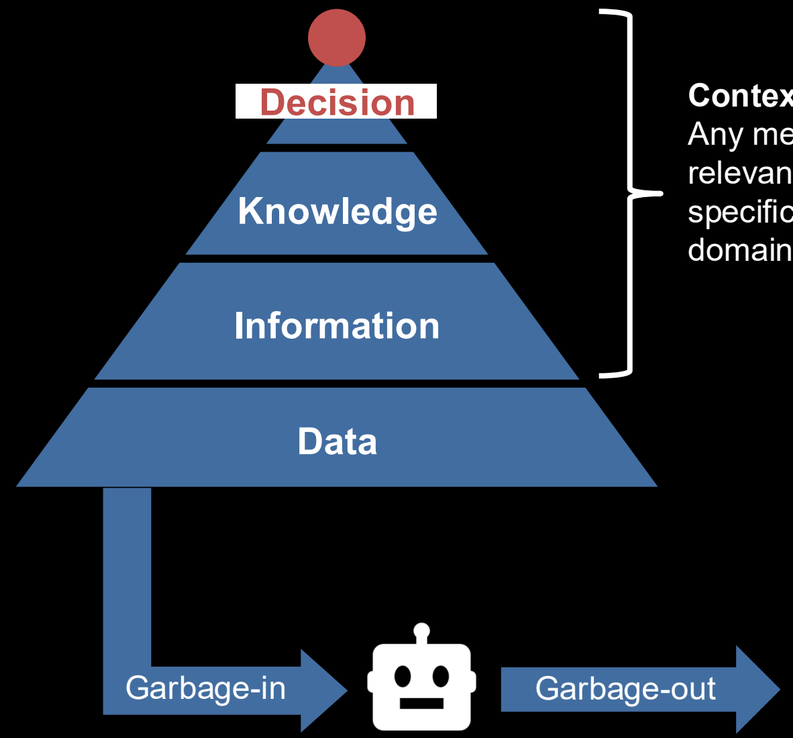
148 MARCH 1988—MANUFACTURING ENGINEERING

Technology-centric mindset

Business Encoding



Data/Context Readiness



https://www.biomedima.org/project/messianism-in-ai-times-keynote/

https://en.wikipedia.org/wiki/DIKW_pyramid

Many Suffer (and will Suffer) from AI

Benefits for few people; harm for the majority (if we continue on the current ungoverned road)

Direct Material Harm

- Toxic metal extraction under inhumane conditions
- “Dematerialization” myth: digital tech still depends on heavy material resources
- Climate conflict: digital growth ~9%/yr vs. –5% needed (Paris)
- Rising CO_2 emissions
- Confirmed military uses of AI
- High energy demand: training one GPT-scale model ≈ a city’s electricity for ~3 days

Indirect Material Harm

- Severe ecological damage in fragile regions
- Strategic dependence on rare metals
- Urban food supply depends on digital systems (GPS, logistics)
- Land devastation and pollution (e.g. Norilsk-nickel mining)
- Heavy metals from landfills and rivers (Sulfuric acid pollution in Zambia; toxic lead in Thai rice fields)
- Resources diverted from food to tech, causing starvation

Direct Immaterial Harm

- Addiction and attention economy
- Exploitative and violent digital labor (click workers)
- Neocolonial extraction and resource predation
- Mass surveillance and fake news
- Implicit support of oppressive supply chains (e.g. cobalt in Congo)
- Cognitive harm to children

Indirect Immaterial Harm

- People disconnect from reality
- Loss of skills via radical monopoly
- Sedentariness and compulsive content use
- Digital fetishism and selective amnesia
- Reinforcement of dominant narratives
- Suppression of truly disruptive ideas although these are most needed in crisis times…

Most of AI Projects fail to Deliver Value

THE ECONOMIC TIMES | Panache

My Watchlist Sign In

Avail Free Trial of ETPrime

English Edition ▾ | Today's ePaper

Home ETPrime Markets Market Data Masterclass ▾ News Industry SME Politics Wealth MF Tech AI Careers Opinion NRI Panache

Luxury Panache ▾ ET Magazine

Business News › Magazines › Panache › MIT study shatters AI hype: 95% of generative AI projects are failing, sparking tech bubble jitters

MIT study shatters AI hype: 95% of generative AI projects are failing, sparking tech bubble jitters

By Paurush Omar, ET Online • Last Updated: Sep 09, 2025, 06:23:00 PM IST

“Promises
(from AI/technology vendors)
only bind people believing
in them”

Our proposal

Life Science (and beyond)

“A methodology for systematic, sovereign, responsible and measurable value creation using state-of-the art AI solutions.”

Three Guiding Principles

1 Harness Digital Complexity

2 Adopt Digital Sobriety and Sovereignty

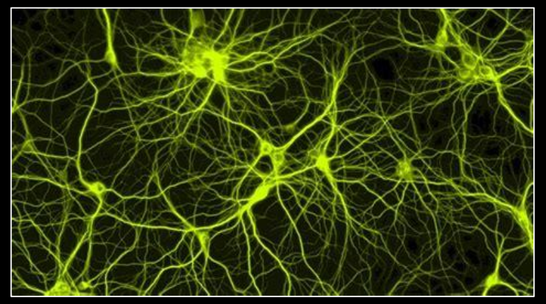
3 Data Governance 4.0 to Ensure AI Added-Value

Proposal

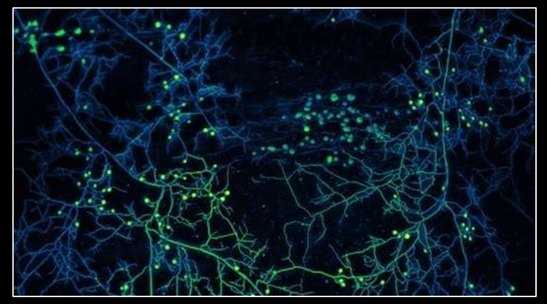
Life Science (and beyond)

1. Harness Digital Complexity

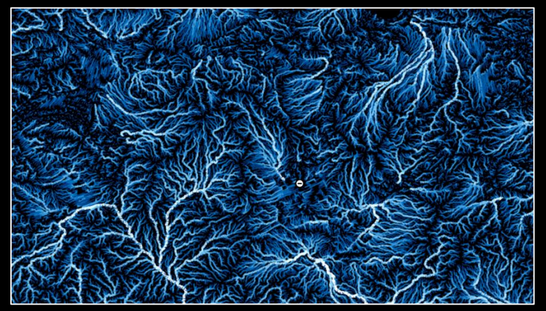
1. Harness Digital Complexity



Human neurons



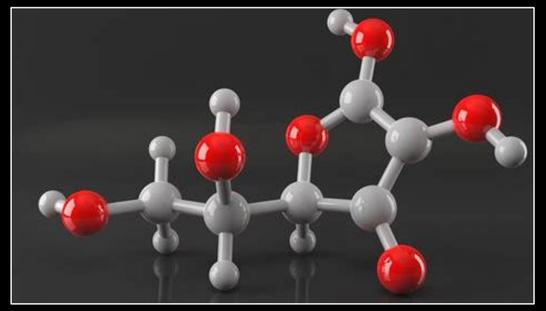
Underground fungi



River basins

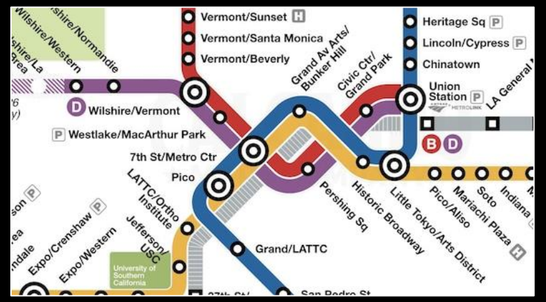
Spider webs

... leaf venation network, plant root system, ant foraging trail, blood vascular system, lung bronchial tracks ...



Molecular model

Constellations

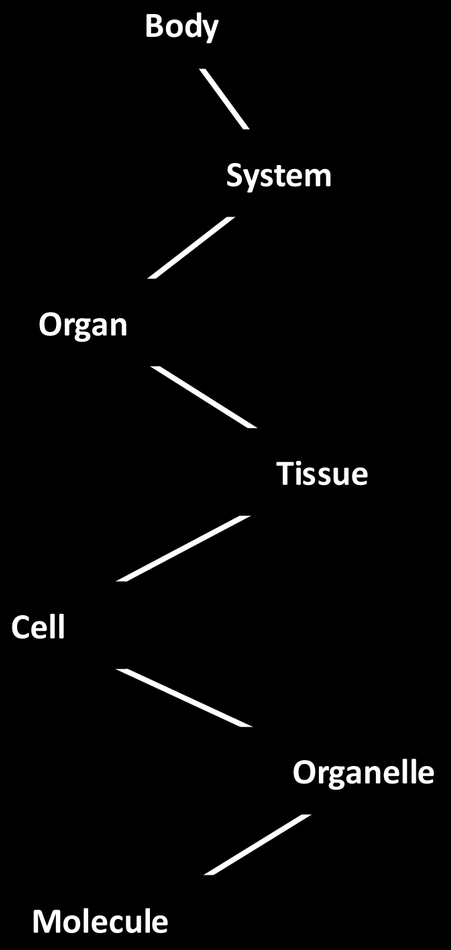


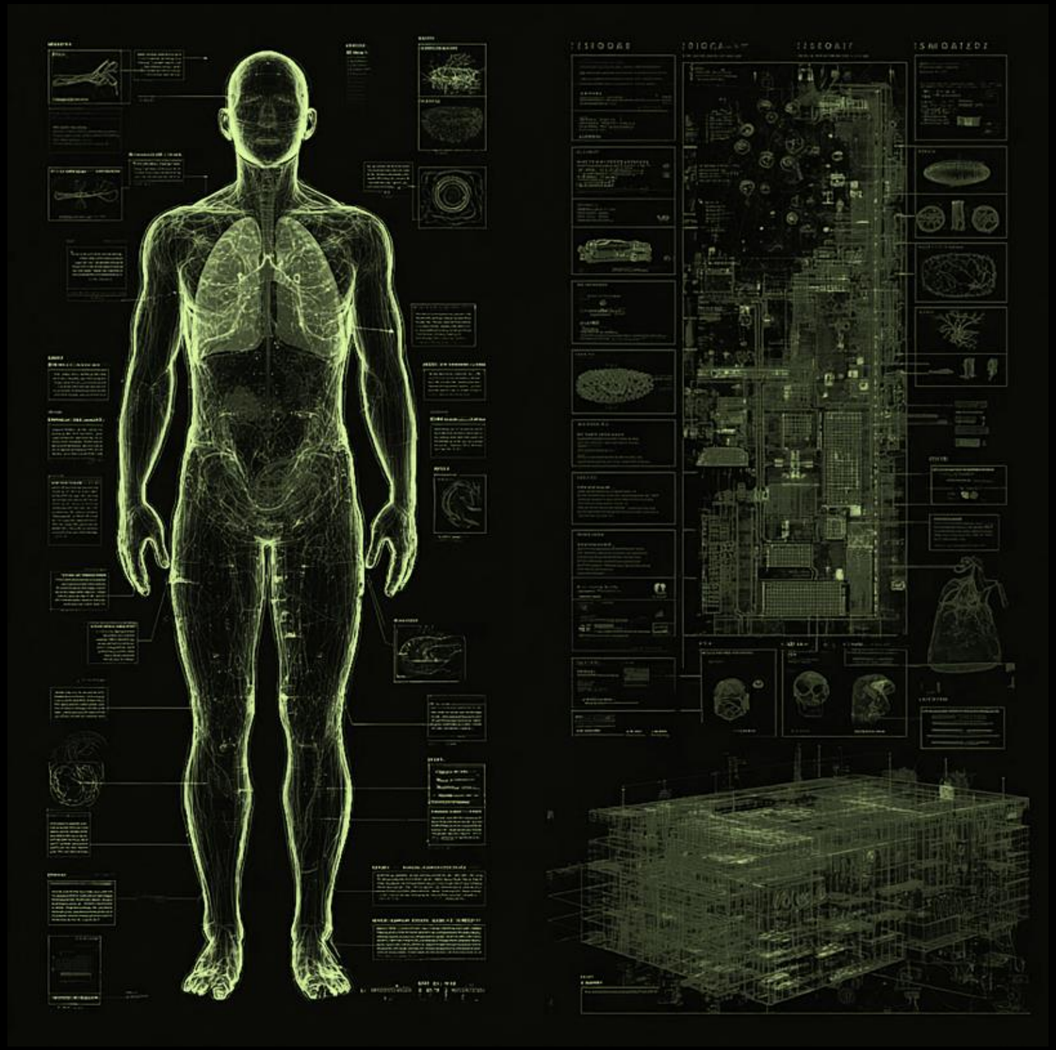
Metro maps

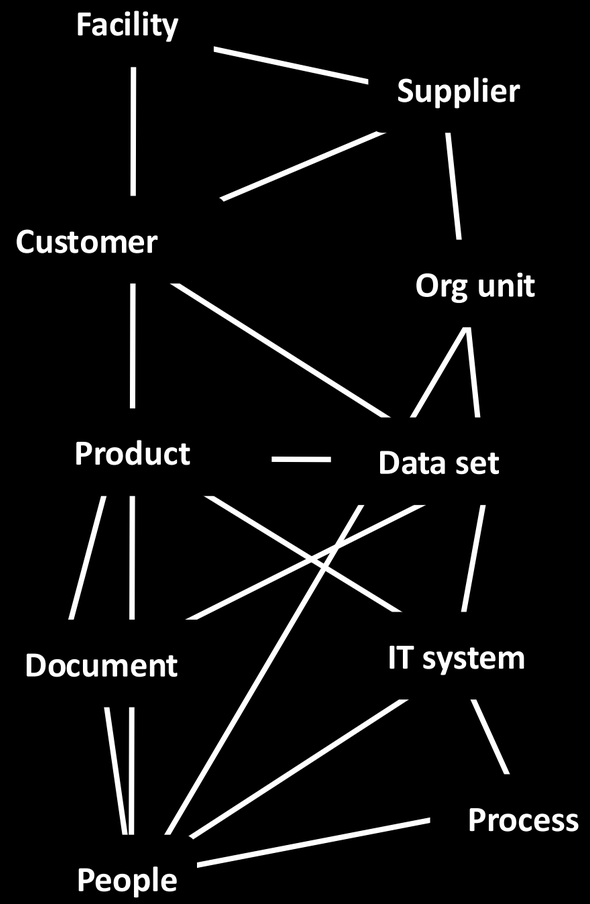
Org charts

... electrical circuits, genealogical diagrams, social network, citation networks in scientific literature, metabolic pathway ...

Knowledge Graph Describe Anything

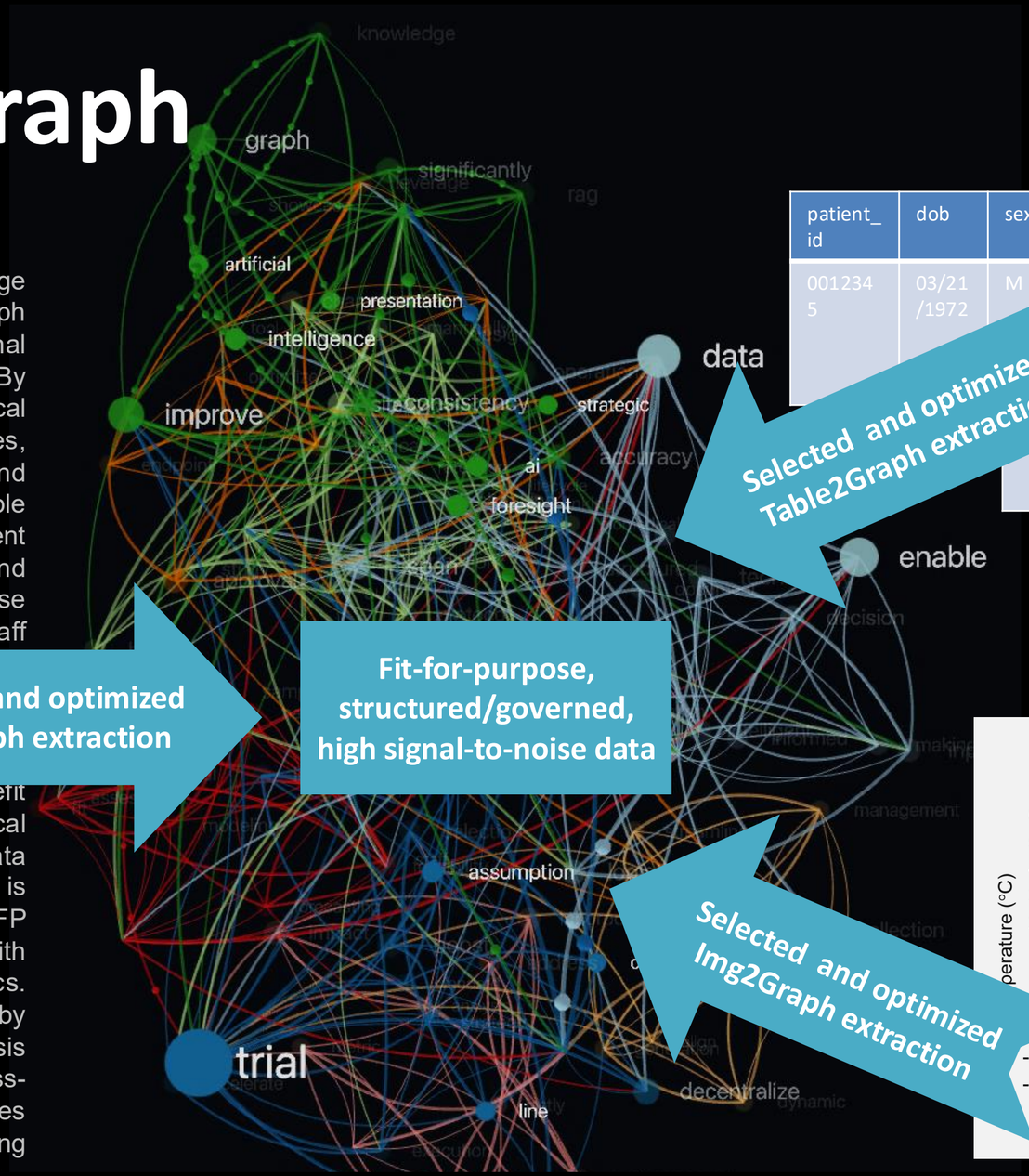






Anything2Graph

This presentation showcases how Knowledge Graphs, Artificial Intelligence (AI), and Graph RAG can significantly improve operational efficiency across the clinical trial lifecycle. By semantically structuring and linking biomedical concepts, protocol elements, site capabilities, patient characteristics, operational and logistical inputs, these technologies enable more informed decision-making and intelligent automation (document writing/checking and structured data transformation). Practical use cases span site feasibility assessments, staff readiness, and procedural
handling and clinical supply
data-enabled decentralized tr
boost patient recruitmen
Protocol authoring and study design benefit from intelligent tools that leverage historical endpoint approvals and optimize data collection frameworks. Budget management is streamlined through automated RFP generation and dynamic tracking aligned with trial assumptions and performance metrics. Lastly, sustainability is addressed by incorporating environmental impact analysis into trial method selection, in line with cross-industry standards. Together, these capabilities accelerate trial execution while improving



patient_id	dob	sex	visit	Site_id	Dia_code	Dia_sys	Dia_desc
0012345	03/21/1972	M	2025-07-14	HOSP-01	I10	ICD-10	Essential (primary) hypertension

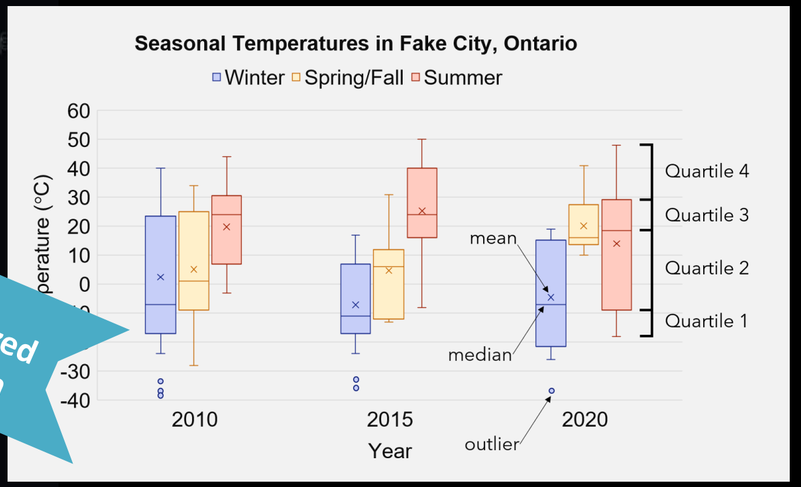
Lab_tst_sys	Lab_tst_nm	Lab_res_val	Lab_res_unit	med	dose
LOINC	Glucose [Mass/volume] in Serum or Plasma	118	mg/dL	Metformin 500 MG Oral Tablet	500 mg

Selected and optimized Table2Graph extraction

Selected and optimized Doc2Graph extraction

Fit-for-purpose, structured/governed, high signal-to-noise data

Selected and optimized Img2Graph extraction

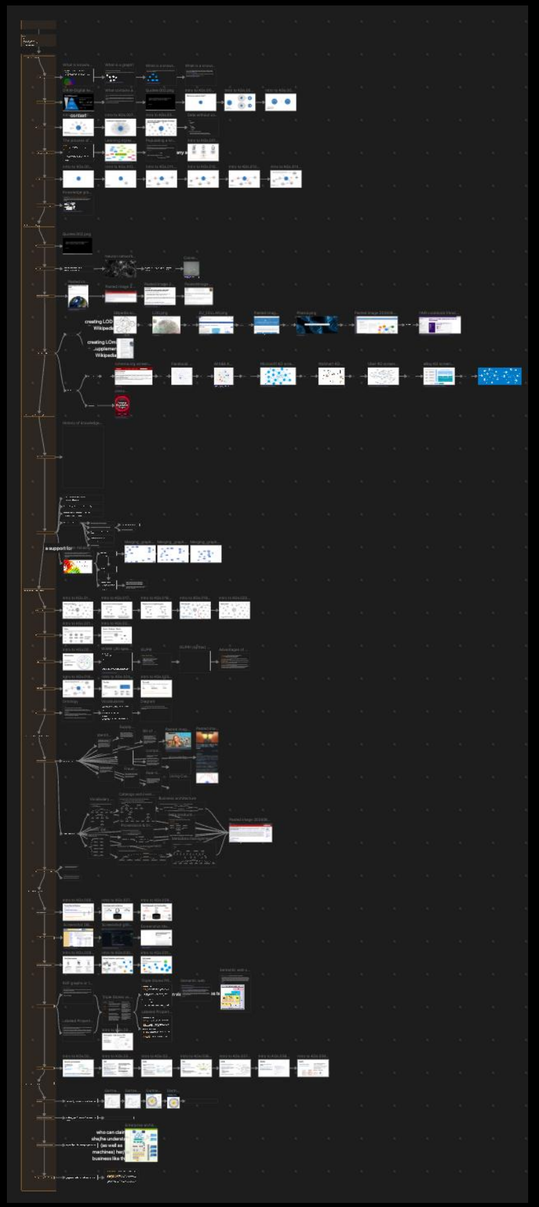


KG have become standard approaches to merge distributed data sources

More about Knowledge Graphs

BIOMED ima

https://www.biomedima.org



https://www.biomedima.org/project/introduction-to-knowledge-graphs/

Proposal

Life Science (and beyond)

2.1 Adopt Digital Sobriety

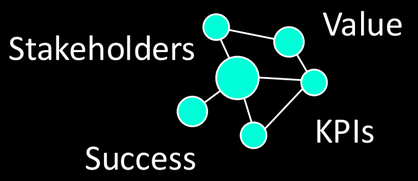
Digital Sobriety

Value depends on context
without context description, no proper value measurement [1]

- *How do you define value?*
- *What is value for you is not necessarily value for me*

Money is a mean, not a purpose !

Graph

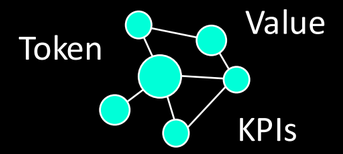


Slow food → Slow business [2]
focus on quality rather than quantity

- *How much you spend on AI tokens?*
- *How each token generate value?*

Lose time to gain time.

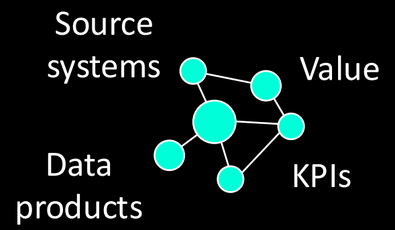
Think first !



Only the data you need !
end-to-end lineage between data sources and success metrics

- *What is the cost of collecting, preparing, storing data?*
- *What value generates your data?*

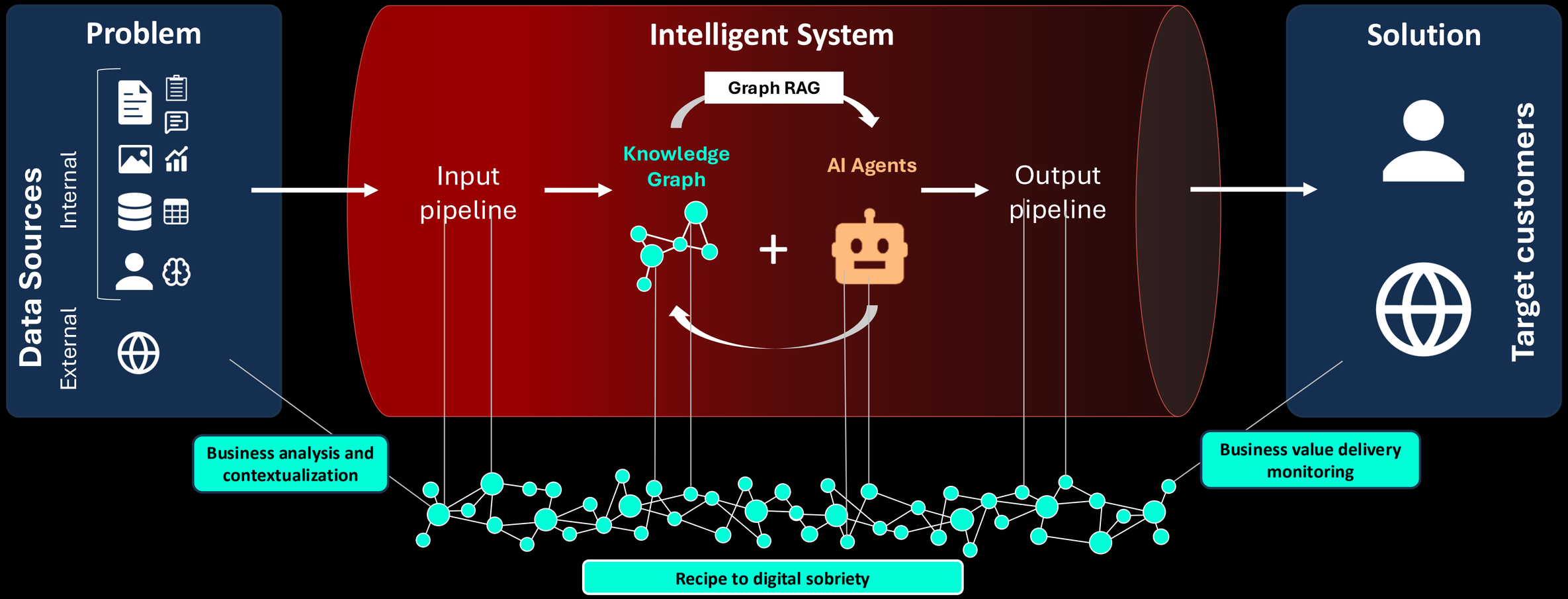
Data is a mean, not a purpose !



1) https://www.biomedima.org/business-value-monitoring/
2) https://www.biomedima.org/slow-business/

Only What you Need to Succeed

Customer-Driven Data collection and processing; end-to-end traceability from requirements to value



Proposal

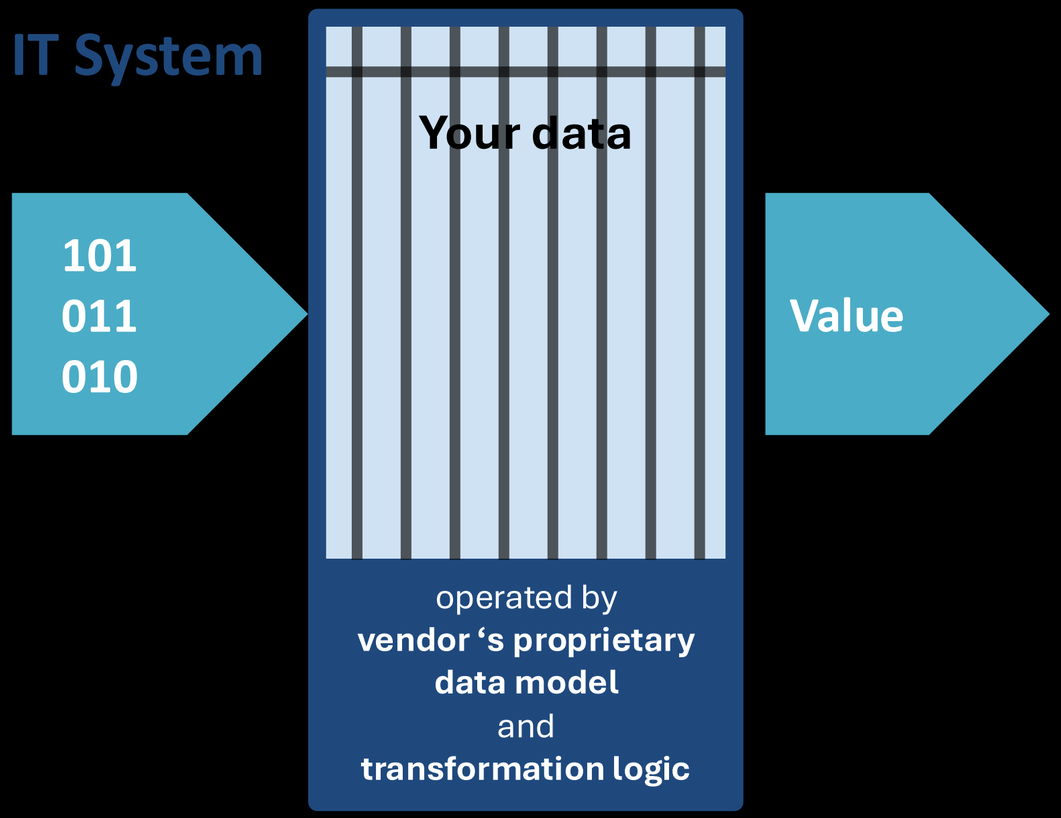
Life Science (and beyond)

2.2 Digital Sovereignty

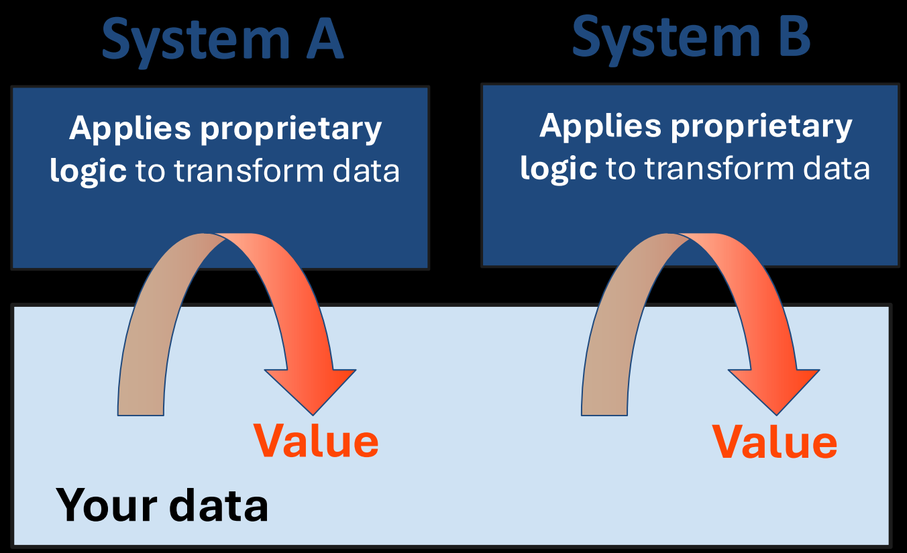
Digital Sovereignty

Avoid technology vendors to lock your data in their system; de-couple your data from their model/logic

Current situation
(system-centric)



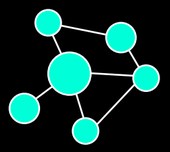
Proposed situation
(data-centric)



Data layer

Metadata layer

Your metadata,
your data model, your data governance logic



Proposal

Life Science (and beyond)

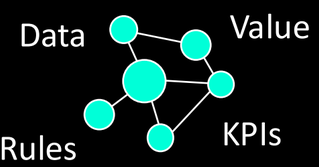
3.1 Data Governance 4.0

Data Governance 4.0

Fit-for-purpose

Governance comes with constraints
For it not to be counter-productive, constraints must not hinder data usage to create value.

Graph

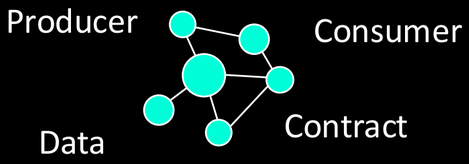


Dynamic

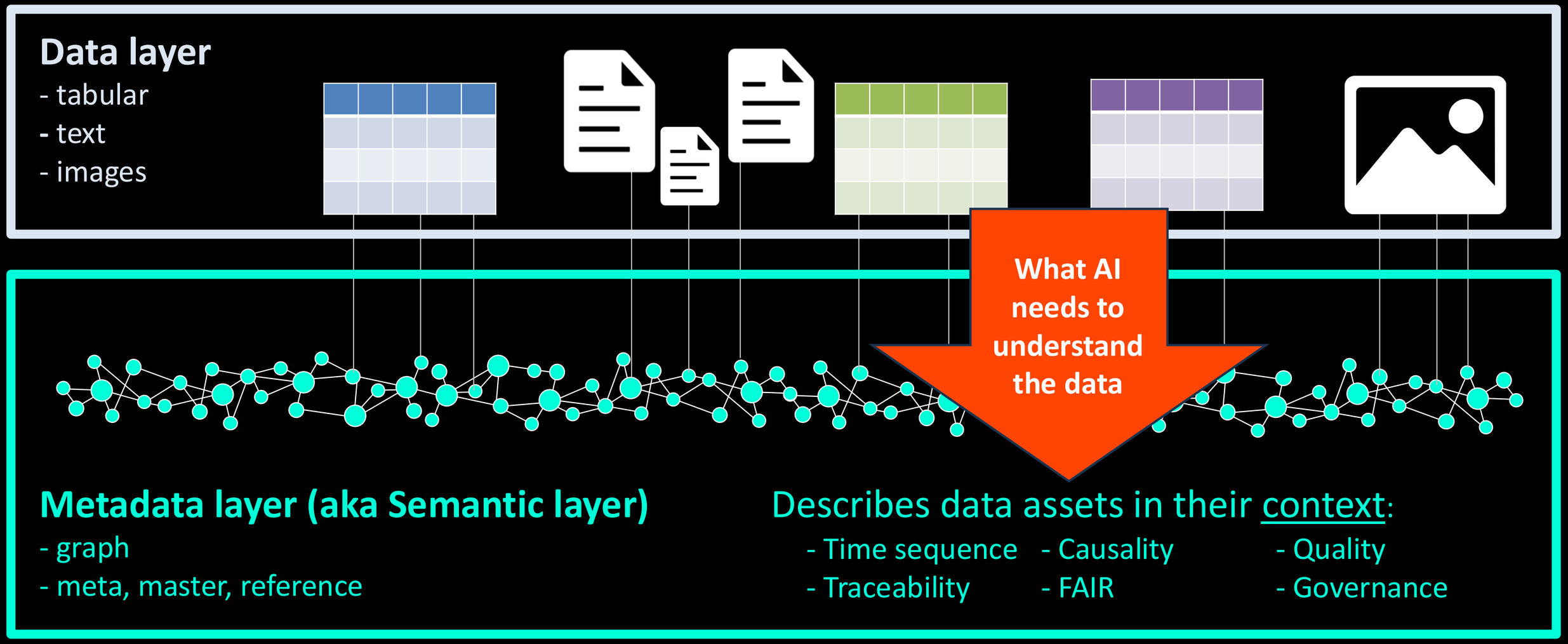
Business changes by the day, how do you adapt?
Governance constraints must adapt business changes to de-risk data but still enable value-generation.

Metadata driven

Data changes frequently, in a decentralized manner
You govern once in metadata and enforce everywhere, consistently and at scale.



Semantic Layer and Context Graph

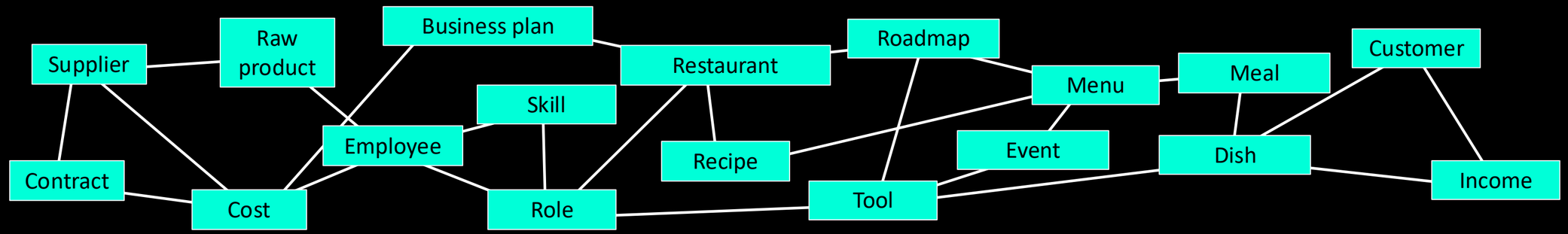


Metadata-Driven Digital Business

Business context




Graph digital twin



Metadata standards

House-keeping models (cross-industry)

Business architecture

Data set
Employee
Facility
Process
IT system

Business process

Activity
Task
Role
Outcome

Data

Data set
Document
Table
Chunk
Column
Value

Business-specific models

Research

Gene
Organ
Protein

Clinical

Disease
Patient
Symptom

Regulatory

Jurisdiction
Market approval

https://arxiv.org/abs/2311.02082

Computer Science > Artificial Intelligence

[Submitted on 20 Oct 2023 (v1), last revised 23 Nov 2023 (this version, v3)]

Semantic Modelling of Organizational Knowledge as a Basis for Enterprise Data Governance 4.0 -- Application to a Unified Clinical Data Model

Miguel AP Oliveira, Stephane Manara, Bruno Molé, Thomas Muller, Aurélien Guillouche, Lysann Hesske, Bruce Jordan, Gilles Hubert, Chinmay Kulkarni, Pralipta Jagdev, Cedric R. Berger

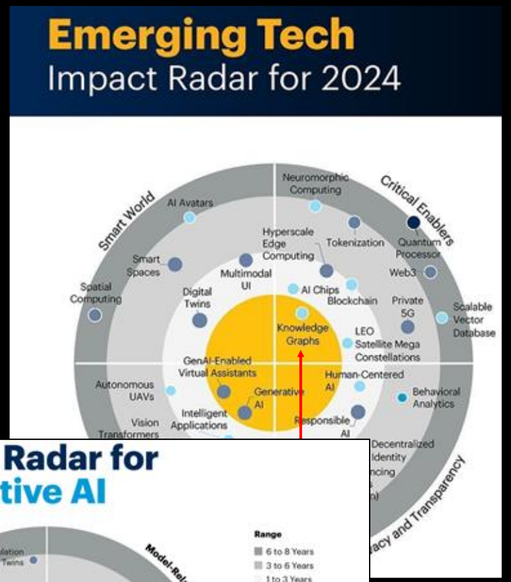
Individuals and organizations cope with an always-growing amount of data, which is heterogeneous in its contents and formats. An adequate data management process yielding data quality and control over its lifecycle is a prerequisite to getting value out of this data and minimizing inherent risks related to multiple usages. Common data governance frameworks rely on people, policies, and processes that fall short of the overwhelming complexity of data. Yet, harnessing this complexity is necessary to achieve high-quality standards. The latter will condition any downstream data usage outcome, including generative artificial intelligence trained on this data. In this paper, we report our concrete experience establishing a simple, cost-efficient framework that enables metadata-driven, agile and (semi-)automated data governance (i.e. Data Governance 4.0). We explain how we implement and use this framework to integrate 25 years of clinical study data at an enterprise scale in a fully productive environment. The framework encompasses both methodologies and technologies leveraging semantic web principles. We built a knowledge graph describing avatars of data assets in their business context, including governance principles. Multiple ontologies articulated by an enterprise upper ontology enable key governance actions such as FAIRification, lifecycle management, definition of roles and responsibilities, lineage across transformations and provenance from source systems. This metadata model is the keystone to data governance 4.0: a semi-automatised data management process that considers the business context in an agile manner to adapt governance constraints to each use case and dynamically tune it based on business changes.

Proposal

Life Science (and beyond)

3.2 Effective and Trusted AI

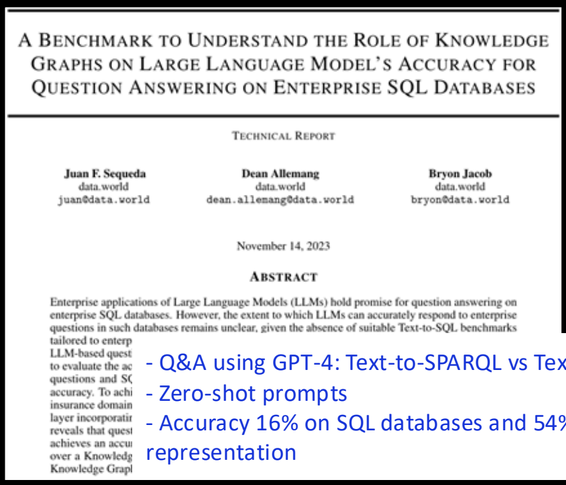
Proven Most Advanced AI Enablers



https://www.gartner.com/en/articles/30-emerging-technologies-that-will-guide-your-business-decisions
https://www.gartner.com/en/articles/understand-and-exploit-gen-ai-with-gartner-s-new-impact-radar

February 13, 2024

https://www.microsoft.com/en-us/research/blog/graphrag-unlocking-llm-discovery-on-narrative-private-data/
https://arxiv.org/html/2404.16130v1



A BENCHMARK TO UNDERSTAND THE ROLE OF KNOWLEDGE GRAPHS ON LARGE LANGUAGE MODEL'S ACCURACY FOR QUESTION ANSWERING ON ENTERPRISE SQL DATABASES

TECHNICAL REPORT

Juan F. Sequeda
data.world
juan@data.world

Dean Allemang
data.world
dean.allemang@data.world

Bryon Jacob
data.world
bryon@data.world

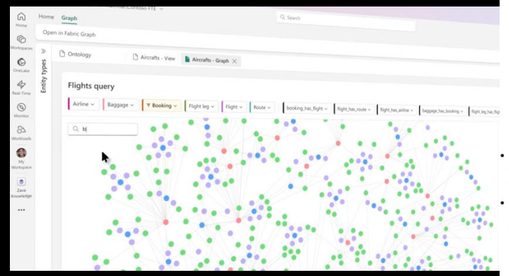
November 14, 2023

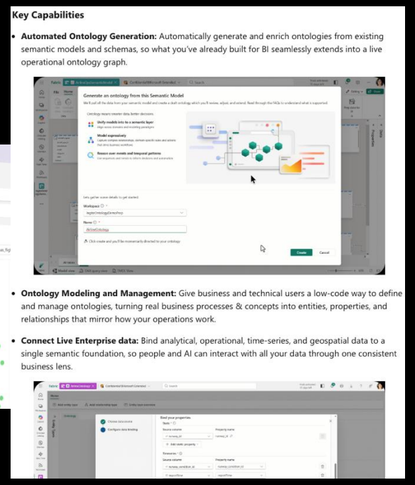
ABSTRACT

Enterprise applications of Large Language Models (LLMs) hold promise for question answering on enterprise SQL databases. However, the extent to which LLMs can accurately respond to enterprise questions in such databases remains unclear, given the absence of suitable Text-to-SQL benchmarks tailored to enterp
LLM-based quest
to evaluate the ac
questions and SC
accuracy. To achi
insurance domain
layer incorporatir
reveals that quest
achieves an accu
over a Knowledg
Knowledge Graph

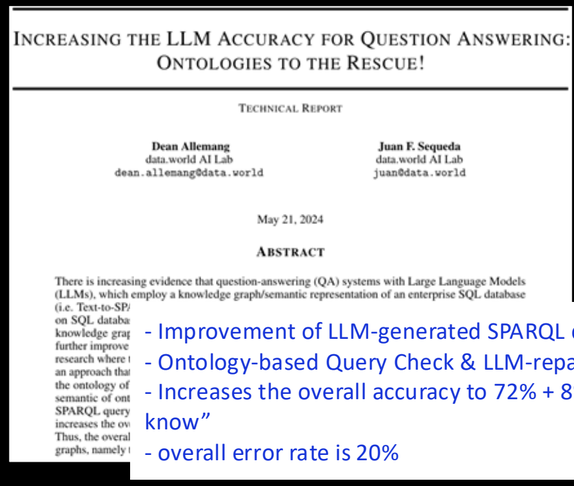
- Q&A using GPT-4: Text-to-SPARQL vs Text-to-SQL
- Zero-shot prompts
- Accuracy 16% on SQL databases and 54% on KG representation

November 19, 2025





https://blog.fabric.microsoft.com/en-us/blog/introducing-fabric-iq-the-semantic-foundation-for-enterprise-ai/



INCREASING THE LLM ACCURACY FOR QUESTION ANSWERING: ONTOLOGIES TO THE RESCUE!

TECHNICAL REPORT

Dean Allemang
data.world AI Lab
dean.allemang@data.world

Juan F. Sequeda
data.world AI Lab
juan@data.world

May 21, 2024

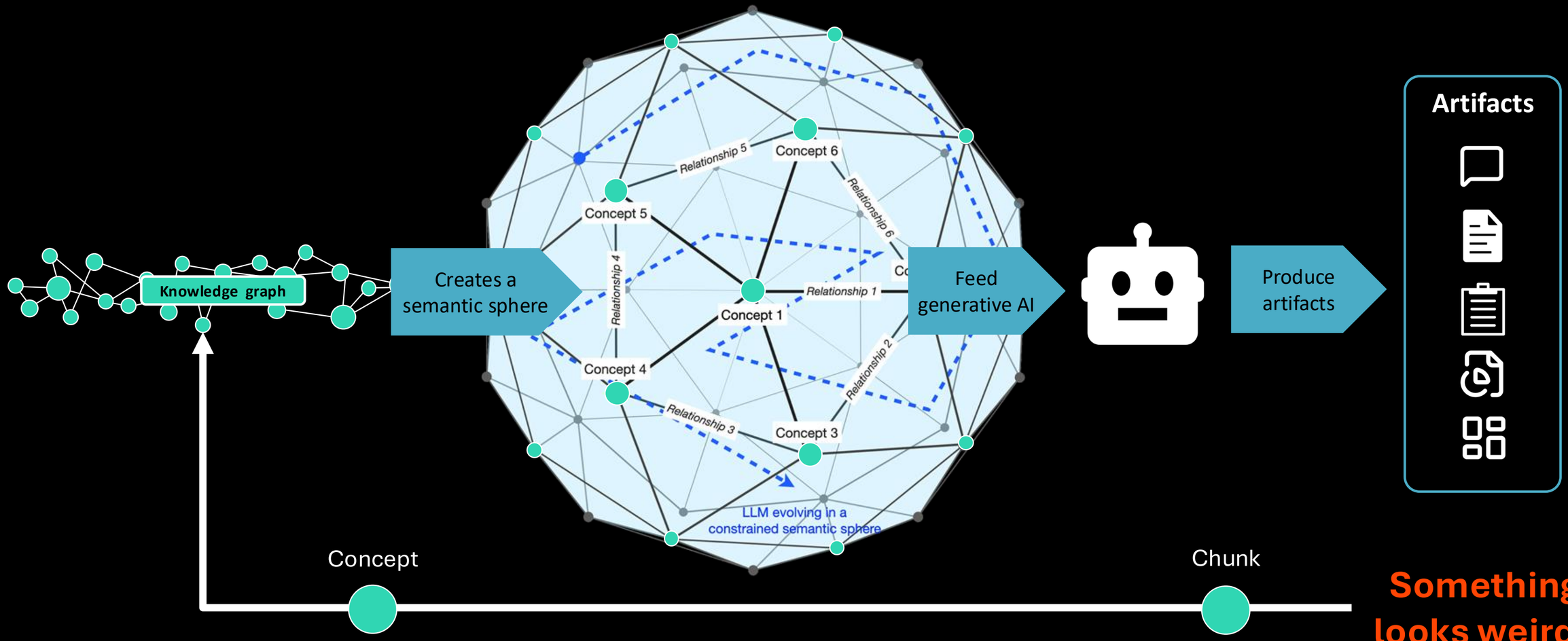
ABSTRACT

There is increasing evidence that question-answering (QA) systems with Large Language Models (LLMs), which employ a knowledge graph/semantic representation of an enterprise SQL database (i.e. Text-to-SP
on SQL databa
knowledge grap
further improve
research where
an approach tha
the ontology of
semantic of ont
SPARQL query
increases the ov
Thus, the overal
graphs, namely

- Improvement of LLM-generated SPARQL queries
- Ontology-based Query Check & LLM-repair
- Increases the overall accuracy to 72% + 8% "I don't know"
- overall error rate is 20%

https://arxiv.org/abs/2311.07509
https://arxiv.org/abs/2405.11706

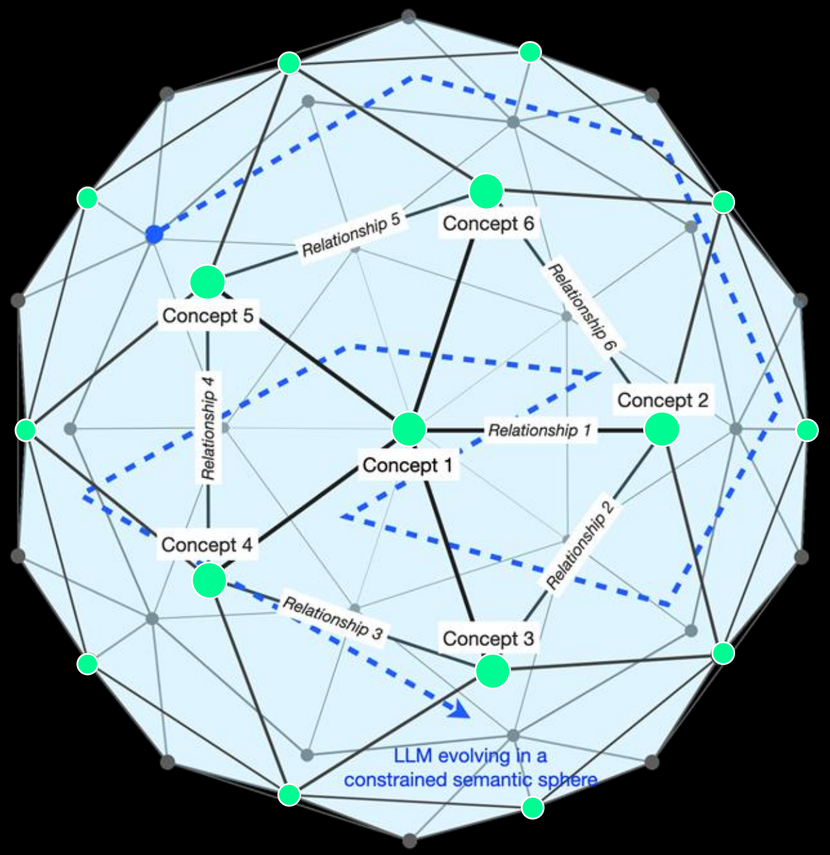
A Closed Semantic Context



Something looks weird?

Traceability to the source (via business architecture model)

Guardrails and Traceability to Ensure Trust



The right context and guardrails to get maximum value out AI

Data/documents

- Client-specific and use case specific
- Mapping of source concepts (documents) to target (eCTD)
- Client-specific document generation

AI

- Restrained hallucination
- End-to-end traceability > transparency and trust in AI output

Rules as graph/code

- Documentation required for system qualification
- Procedure to follow for system validation

Conclusion

Our Proposal to Get Value out of AI

Harness VUCA complexity

→ **Graph is the way**

Relational DBs are good for key-value pairs;
intertwined connections require graph

There is no one-size-fits-all

→ **Personalized medicine**

Your use case is unique; your context is unique
You need a unique setup to deliver value

Data Governance 4.0

- Fit-for-purpose
- Dynamic
- Metadata-driven

AI to develop vs. AI to assist people

→ **Knowledge Management (KM)**

Without KM, AI accelerates forgetting;
with KM, AI accelerates learning.

Thanks for your Attention

which is not all we need but a rare, hence precious thing nowadays

https://www.linkedin.com/in/cedricberg/

www.biomedima.org

www.migx.ch