

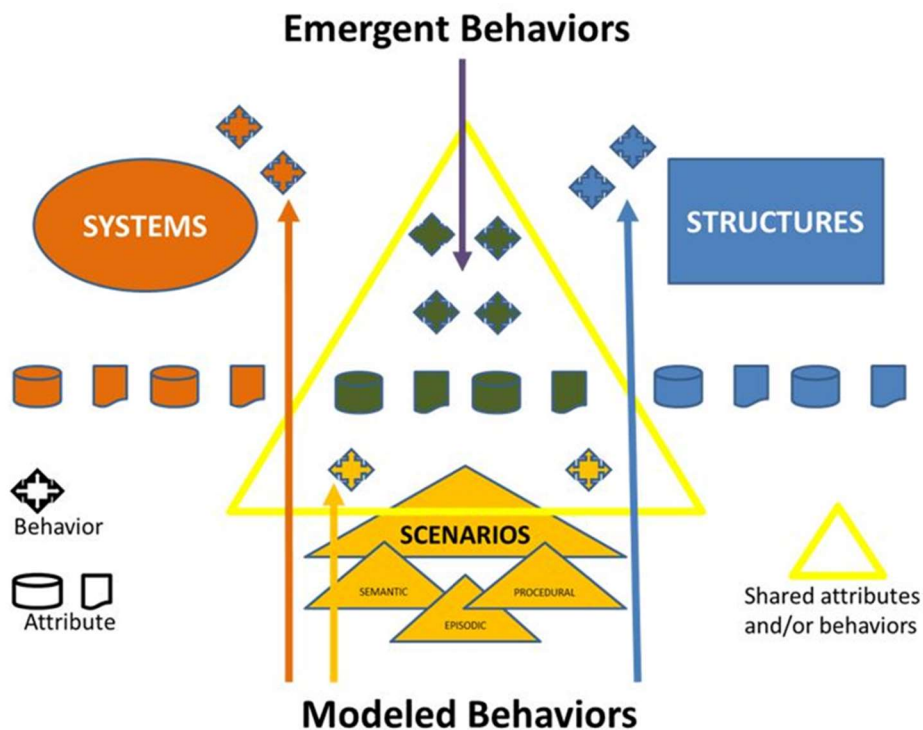
RYALITI Dark Data Discovery Innovation Overview

Knowledge Engineering using a Biomimetic Engine

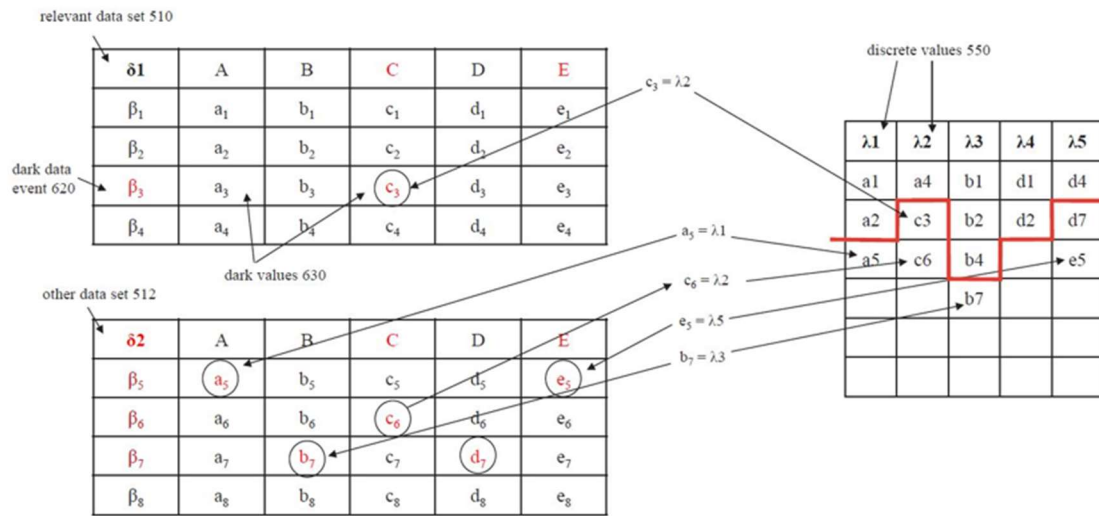
On March 3, 2022, the National Academy of Science (NAS) published a report on The Physics of Life- (<http://nap.edu/26403>), with the following conclusion on the future of AI, “An important lesson from the long and complex history of neural networks and artificial intelligence is that revolutionary technology can be based on ideas and principles drawn from an understanding of life, rather than on direct harnessing of life’s mechanisms or hardware. Although it may take decades, it thus is reasonable to expect that principles being discovered today will inform the technologies of tomorrow.” We reached that conclusion some years ago and applied it.

A Biomimetic Engine is an ecosystem of independent digital twins designed using principles and methods discerned from the human brain. The primary applications of the engine are:

- Practical modeling of highly complex, multi-disciplinary domains to enable holistic discovery and analytics to find realities that are invisible to current ML/NLP, e.g., multi-factor commonality, dark data (Gartner), latent variables.
- Chaotic and stochastic simulation that is traceable and repeatable.
- Capture elements of human expertise as a computable asset that can be leveraged by standard architectures via an API.
- The ecosystem models structures, systems and scenarios as illustrated here:



The result of this approach is a toolset that not only facilitates agile design, but also dynamic simulation, exploration, and experimentation which includes dark data, as illustrated here:



This is an engineering-level view of dark data, which cannot be seen by traditional applications or AI because:

1. Algorithms can only find what they have been engineered to find.
2. AI can only work within the narrow limits of the algorithm's training data.

In this case multifactor correlations between pathogenic variants (relevant), VUSs (other) and knowledge graphs populated by the researchers produced the findings reported here.

Ecosystem Architecture

- Each twin models a discrete component of the analytical scope of the ecosystem.
- Internal properties and behaviors must be modeled to a level of sufficient comprehensiveness to enable the reactions that are required for the ecosystem to reflect the real world to the scope of its design.
- Each twin can initiate an interaction with others or respond as prompted.
- Mitigation of bias is achieved by:
 - Independent design of each twin
 - Abstract knowledge graphs populated without defining specific problems or events
 - Autonomous interactions between the twins

This real-world reasoning approach enables the construction of models that integrate highly diverse elements and information sources to enable exploration and discovery to a scope that traditional information architecture cannot accommodate.

Systems Thinking and Real-world Reasoning (RWR)

The NAS also recommends addressing complexity using systems thinking. Key observations are:

- Bottom-up, mechanistic, linear approaches to understanding macro-level behavior are limited when considering complex systems
- Bottom-up, reductionist hypotheses and approaches can lead to a proliferation of parameters; this challenge can potentially be addressed by applying top-down, system-level principles
- Systems thinking can be used to predict macroscopic phenomena while bypassing the need to explicitly unmask all the quantitative dynamics operating at the microscopic level

While all knowledge engineering efforts seek to incorporate elements of cognitive science, a key aspect of the RYAILITI innovation strategy is the driving role of a cognitive methodology, which is enabled by biomimetic information architectures. Brain processes are SYSTEMIC and leverage what neuroscientists label PLASTISITY and SPARSITY.

- Plasticity is the ability to engage diverse combinations of neurons and synapses by relevance to the purpose of the analysis, and to dynamically adapt internal functional architectures.
- Sparsity is the ability to identify the minimum data required. The brain can respond to situations that are simultaneously new on multiple dimensions and can even categorize one data point.

The neuronal and synaptic architecture of the brain is an ecosystem, which according to the National Academies of Science contains 100 trillion neurons. Systemic architecture, plasticity and sparsity are core to biological learning, but are NOT similar to ML algorithms. The biomimetic technologies that enable elements of RWR are:

- EXPERTESE GRAPHS and
- NEURAL SYSTEM DYNAMICS DIGITAL TWINS

We can imitate principles of plasticity and sparsity by implementing qualitative expertise graphs and leveraging them for contextual selection of data and methods from the in-memory model library. Unlike the deterministic methods to which traditional application engineering is limited of necessity, systemic modeling requires the coexistence of chaotic and stochastic model elements, as well as their ability to dynamically interact with the deterministic elements.

For several years AI has been looked to as the leading pathway to genetic understanding and drug development (<https://doi.org/10.1155/2018/6217812>), deep learning (DL) and natural language processing (NLP) have three key challenges that are addressed by the biomimetic digital twin ecosystem methodology presented in this article;

1. Instability
2. Blindness to dark data
3. Risks and biases that are being challenged by the FDA

Architecting Stability

“Instability is the Achilles’ heel of modern artificial intelligence.” (NAS March 16, 2022) This vulnerability grows as neural networks scale. The historical approach has been mathematical, but reality is dynamic and multidimensional, and emulation requires multidisciplinary methods, at the very least including neuroscience, linguistics, biophysics, and psychology.

Is brain stability a challenge? Neuroscientist Henry Markram, director of the Blue Brain Project (building a computer simulation of a mouse brain – 2020 interview) describes the mathematical challenge: “The human brain uses over 20,000 genes, more than 100,000 different proteins, more than a trillion molecules in a single cell, nearly 100 billion neurons, up to 1,000 trillion synapses and over 800 different brain regions to achieve its remarkable capabilities. Each of these elements has numerous rules of interaction. Now it would take a very long time to even try to calculate how many combinations there are when one can choose 10,000 possible genes from a set of 20,000, but what is sure is that it is more than the number of sub-atomic particles in this, and probably every other universe we can imagine.”

How is stability achieved in the human brain? Markram explains: “Every biological parameter in the brain is constrained by virtually every other parameter in the brain. The parameters can therefore not just assume any value, as every other parameter is holding it in place.”

The knowledge engineering platform methodology includes processes and methods that impose internal constraints on parameter values across the network but delivers critical benefits in addition to stability. The constraint architecture then becomes a foundation for governing design and audit processes. It also reduces the combinatorial complexity of the network to make bias analysis feasible.

This approach is a paradigm shift, moving from obscuring reality by narrow technological constraints to discovering reality by applying real-world reasoning to real-world data. But as Einstein said, “You cannot solve problems using the thinking that caused them.”

Dark Data and AI Blindness

The limitations of standard AI have been identified by various expert sources:

Causality has been highlighted as a major research challenge for AI systems. – National Academy of Sciences 2022

Common sense reasoning, (qualitative analysis; minimum data sets; minimum processing methods) – DARPA AI investment priorities defined for the next ten years (MIT Tech Review 4/19)

Premise Detection and Validation – Mainstream neuroscience relies on taking very specific methods and results and packaging them in a vague cloud of concepts that are only barely agreed upon by the field. In a lot of neuroscience, the premises remain unexamined, but everything else is impeccable. - John Krakauer, neuroscientist at Johns Hopkins University

Gartner coined the phrase “dark data” to describe organizational information assets that are excluded from analytical processes, leaving the associated insights invisible and the value unharvested. Addressing each of the three above limitations requires finding and connecting dark data. Standard AI is blind to the dark data because:

1. Algorithms can only find what they have been engineered to find.
2. AI can only work within the narrow limits of the algorithm’s training data.

The strategies to address these limitations are:

Small and Wide Data - Wide data allows analysts to examine and combine a variety of small and large, unstructured and structured data, while small data is focused on applying analytical techniques that look for useful information within small, individual sets of data. (Gartner Says 70% of Organizations Will Shift Their Focus From Big to Small and Wide Data By 2025 – May 2021)

Integrating Human Expertise – The next move should be to develop machine learning tools that combine data with available scientific knowledge. We have a lot of knowledge that resides in the human skull which is not utilized. - Judea Pearl, UCLA-leading theorist on causal reasoning

Biomimetic Abstractions – An important lesson from the long and complex history of neural networks and artificial intelligence is that revolutionary technology can be based on ideas and principles drawn from an understanding of life, rather than on direct harnessing of life’s mechanisms or hardware. – National Academies of Science Physics of Life Report (<http://nap.edu/26403>)

The biomimetic digital twin ecosystem architecture enables three core use cases:

1. Modeling complex options and tradeoffs dynamically, reporting the associated evidence
2. Modeling theories and testing them against the evidence in the lake and outside
3. Exploration and discovery across diverse and complex data

Use cases 1 and 2 leverage dark data to extend the domains of the analytical functions. Use case 3 outputs discovered dark data in potentially relevant context scenarios.