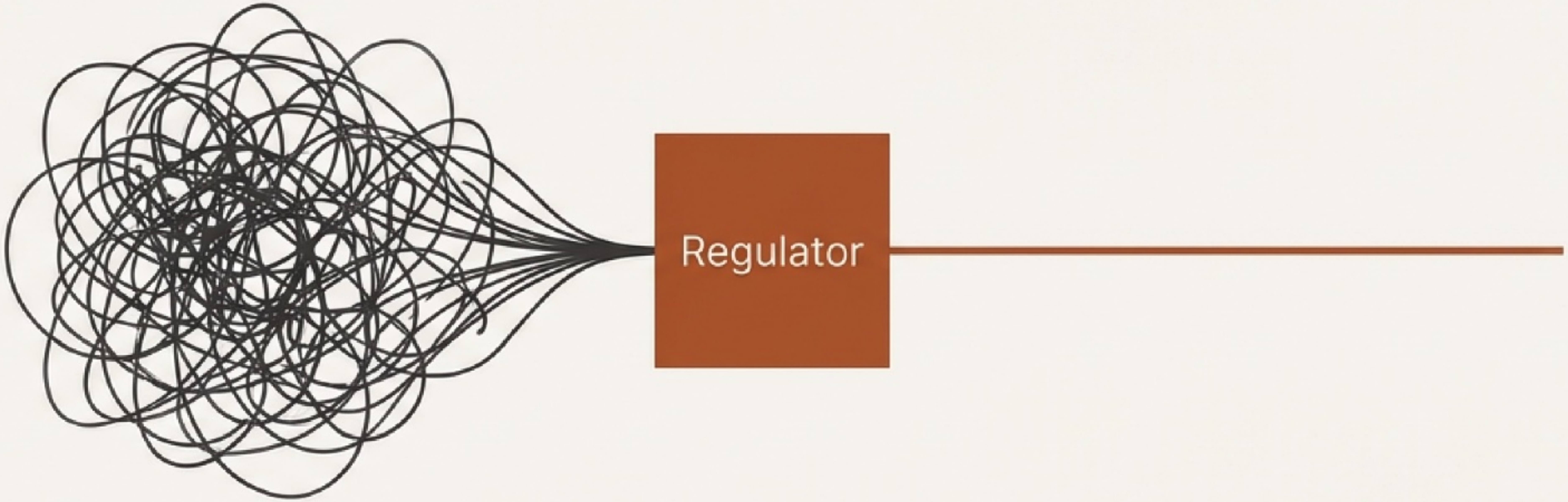
The Algorithmic Regulator

Can we detect an algorithmic agent?



How Simplicity Reveals a Model at Work

Giulio Ruffini

Dec 19 2025

Kolmogorov complexity ($\mathcal{K}$)

Agents need in the soup need to *model* the "world" (Regulator theorem).

But what is a model of a dataset? A short description of the dataset.

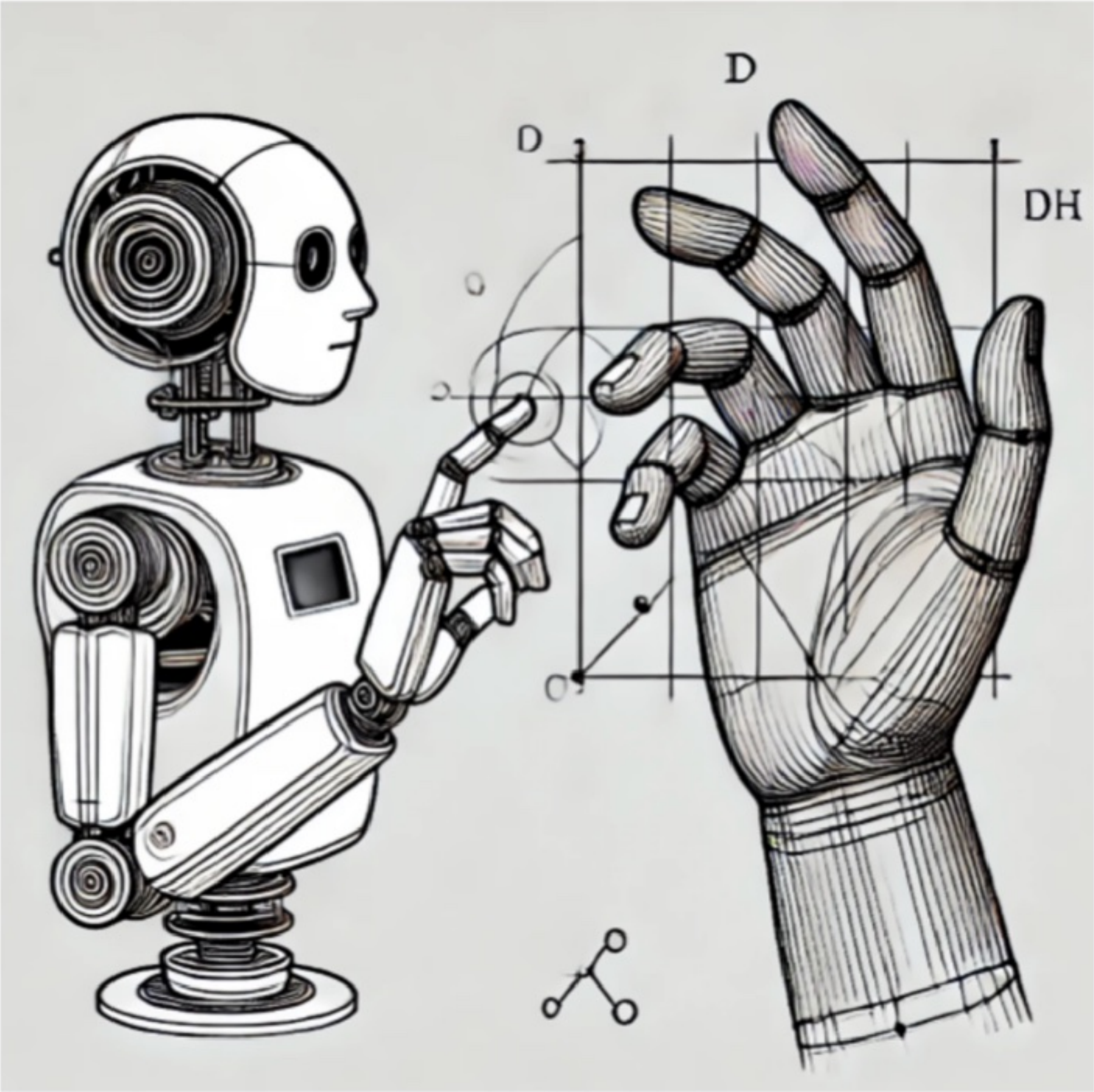
Definition (Model of a dataset)

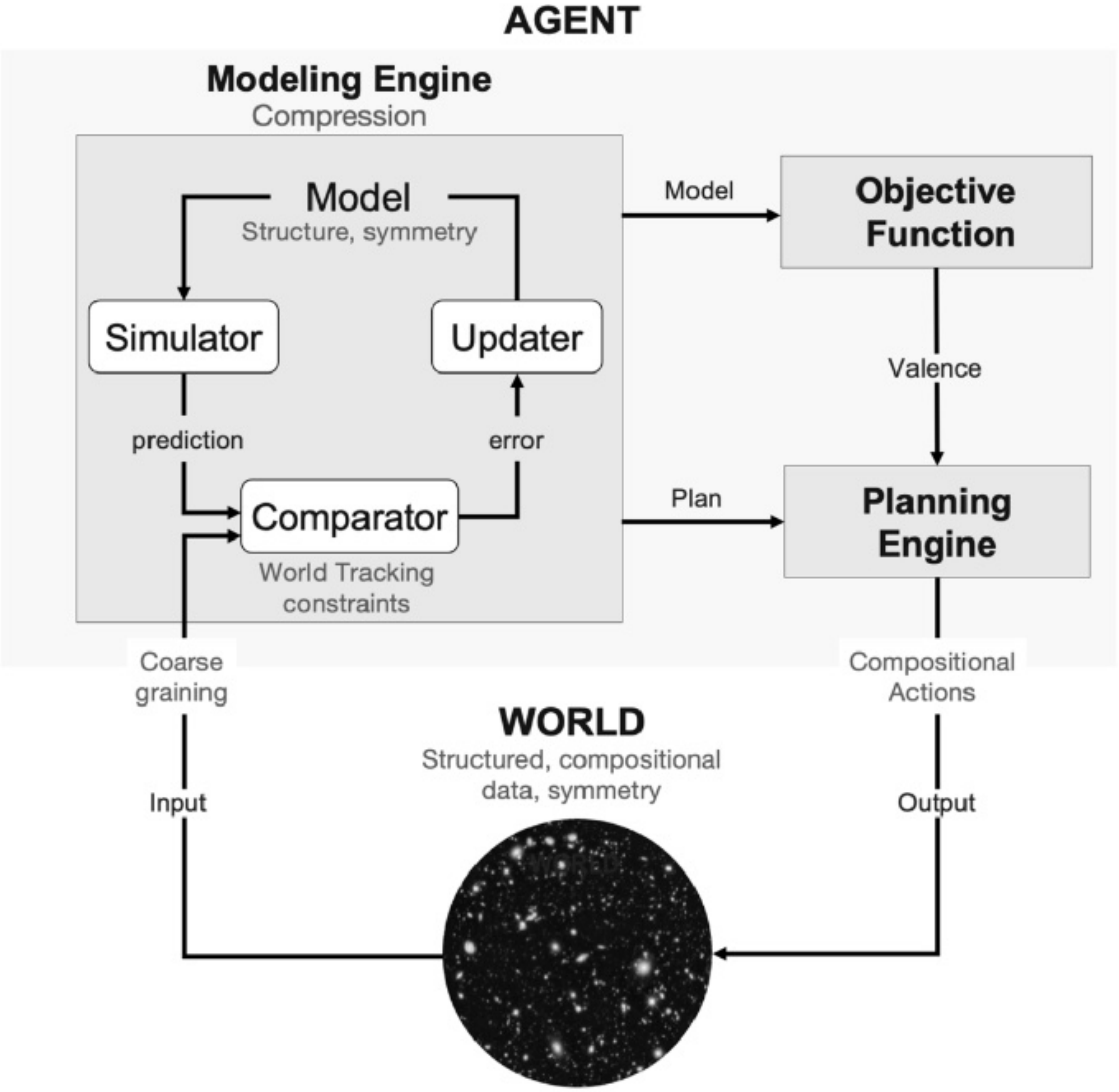
A (succinct) program that generates (or **compresses**) the dataset.

The computational perspective leads us directly into the heart of AIT: the **Kolmogorov complexity** of a dataset ($\mathcal{K}$) is the length of the shortest program capable of generating the dataset[10].

BCOM

The algorithmic agent (minimal model?)



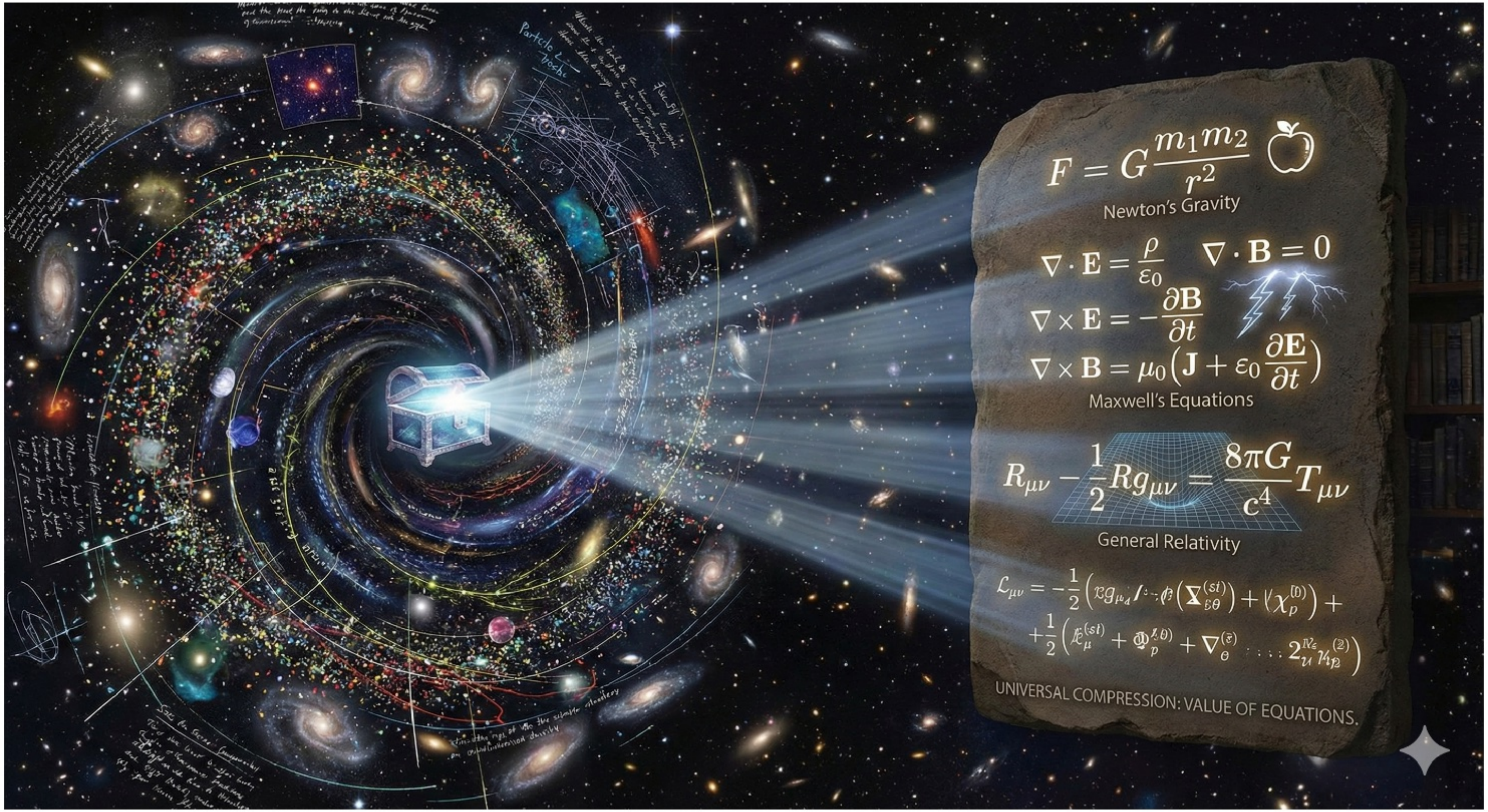


Open Access Perspective

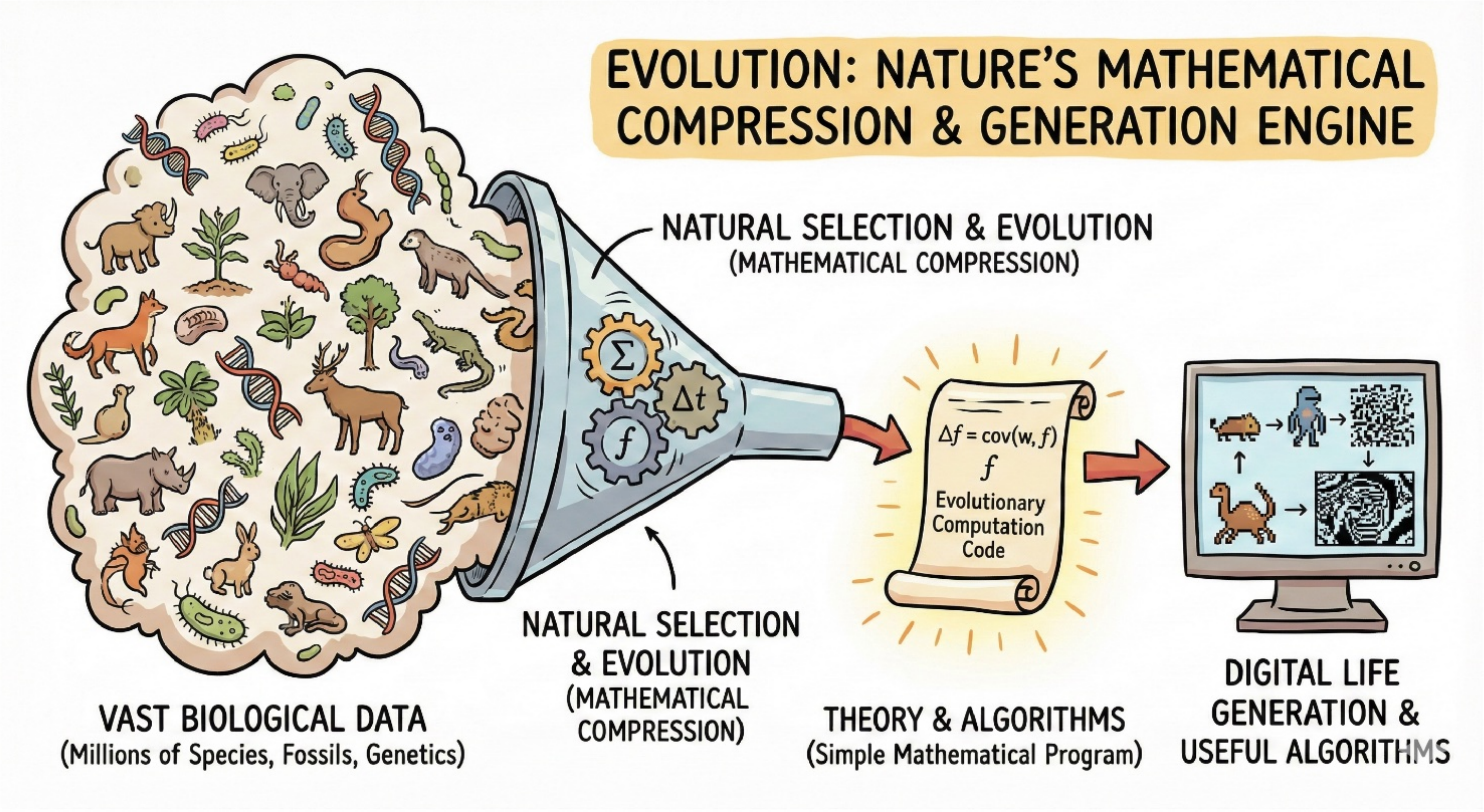
The Algorithmic Agent Perspective and Computational Neuropsychiatry: From Etiology to Advanced Therapy in Major Depressive Disorder

by Giulio Ruffini [1,*], Francesca Castaldo [1,*], Edmundo Lopez-Sola [1,2], Roser Sanchez-Todo [1,2] and Jakub Vohryzek [2,3]

Science as Compression — Physics



Natural Selection as Mathematics



BCOM

Mutual algorithmic information ($\mathcal{M}$)

With $\mathcal{K}$ at hand, we can define an algorithmic version of mutual information:

Definition (Mutual algorithmic information complexity $\mathcal{M}$)

The *mutual algorithmic information* $\mathcal{M}(x:y)$ *between two strings* x *and* y*, is given by*

$$\mathcal{M}(x:y) = \mathcal{K}(x) + \mathcal{K}(y) - \mathcal{K}(x,y)$$

11;12.

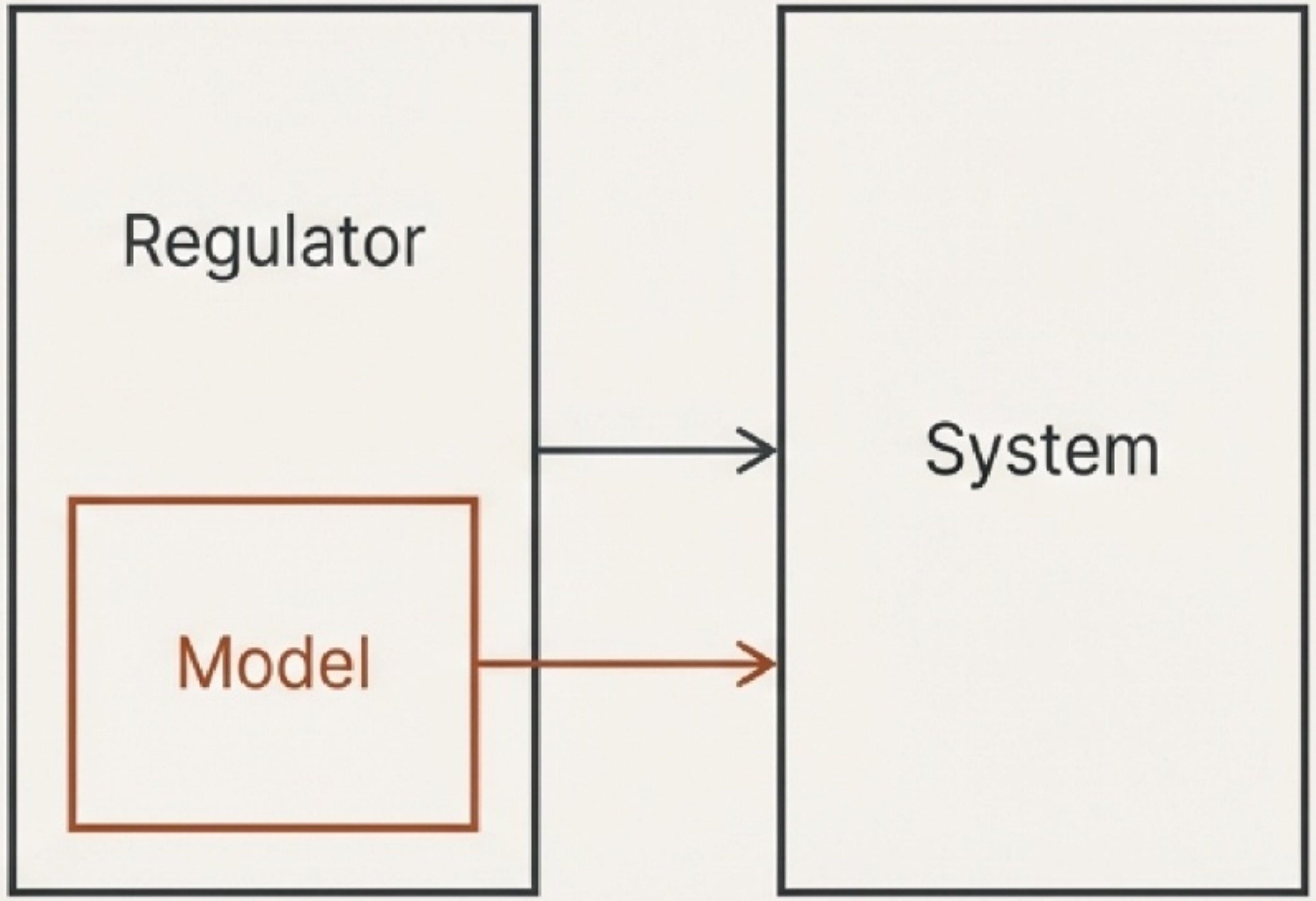
BCOM

The Foundational Intuition

“Every good regulator of a system must be a model of that system.”

—Conant & Ashby, 1970

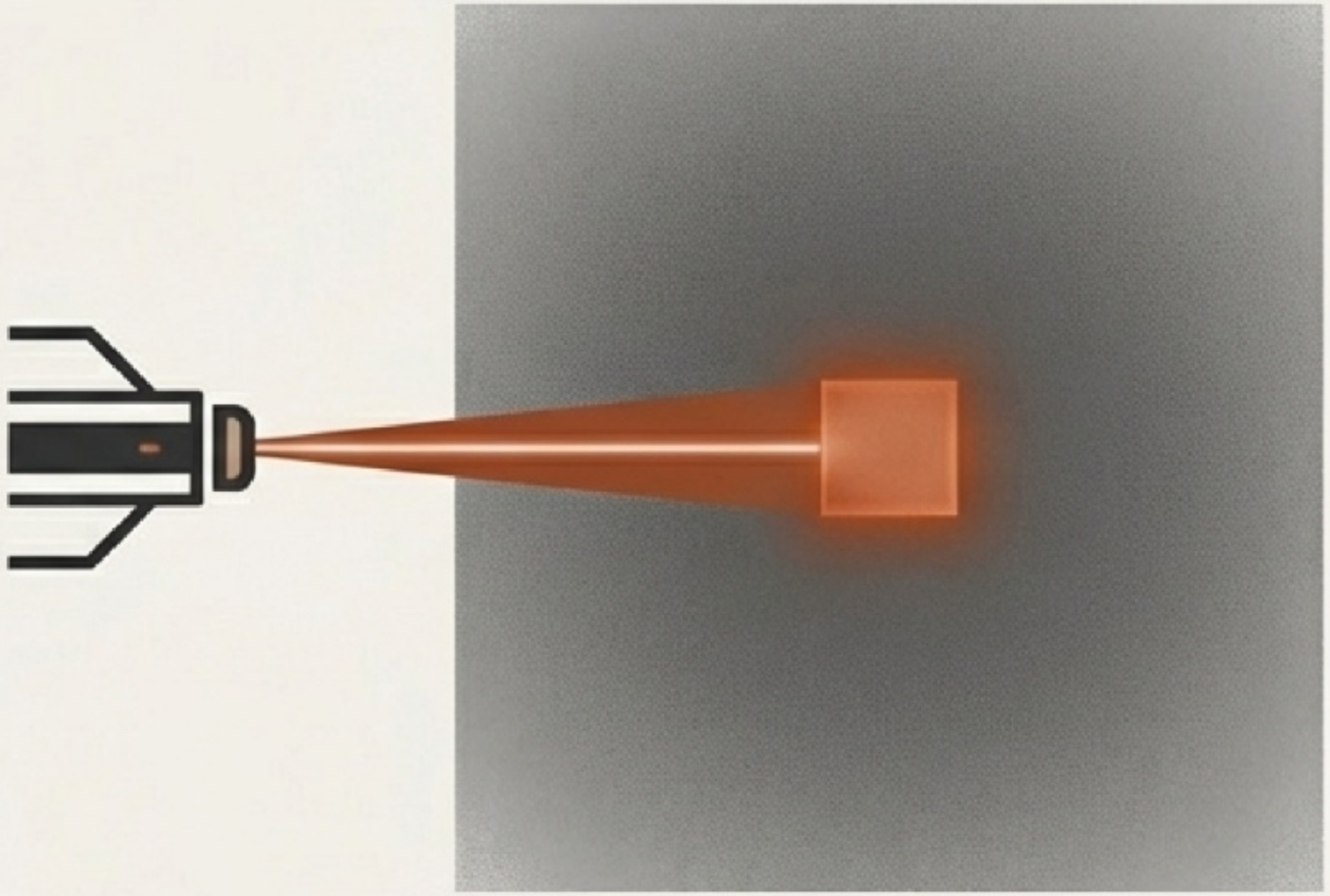
- **Influential**: Underpins core ideas in neuroscience and control theory.
- **Intuitive**: Aligns with our common-sense understanding of control.
- **Informal**: The original theorem has been criticized for vague definitions of “model” and “goodness,” and for a proof that doesn’t fully deliver the headline claim.



The Rigorous, But Restricted, Successor

The Internal Model Principle (IMP) provides the mathematical rigor the Good Regulator Theorem lacked.

- **Precise & Falsifiable**: For a given signal class, it states that a controller must embed a dynamical copy of the signal generator for perfect regulation.
- **Powerful**: A standard backbone for modern robust control.
- **Limited**: The classical IMP is a linear result. It applies to Linear, Time-Invariant (LTI) systems. While non-linear generalizations exist, they require strong structural hypotheses and are not universally applicable.



A New Perspective: Regulation as Compression

We can reframe the problem using Algorithmic Information Theory (AIT). Instead of thinking about “goodness,” we think about **simplicity** and **compressibility**.

A good regulator is one that makes the world's output simple, predictable, and thus, highly compressible.

A) Regulator OFF

B) Regulator ON

High Algorithmic Complexity.

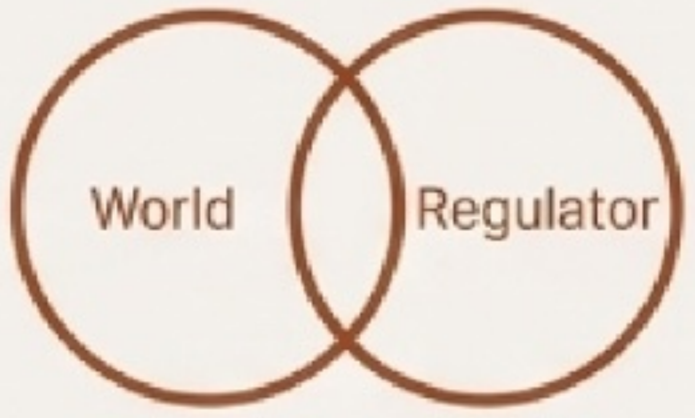
Low Algorithmic Complexity.

The Language of Algorithmic Information

Kolmogorov Complexity, $K(x)$

The length of the shortest program (or recipe) on a universal computer that can produce the string x.

Intuition: It's the ultimate measure of compressibility. A random-looking string has $K(x) \approx |x|$, while a simple string (like "000...0') has a very small $K(x)$.



Mutual Algorithmic Information, $M(W{:}R)$

The amount of shared algorithmic structure. $M(W{:}R) = K(W) - K(W|R)$.

Intuition: The number of bits saved when describing the World W if you are already given the Regulator R. This is our new, rigorous definition of a "model." A regulator "models" the world if $M(W{:}R) > 0$.

Contrastive Testing

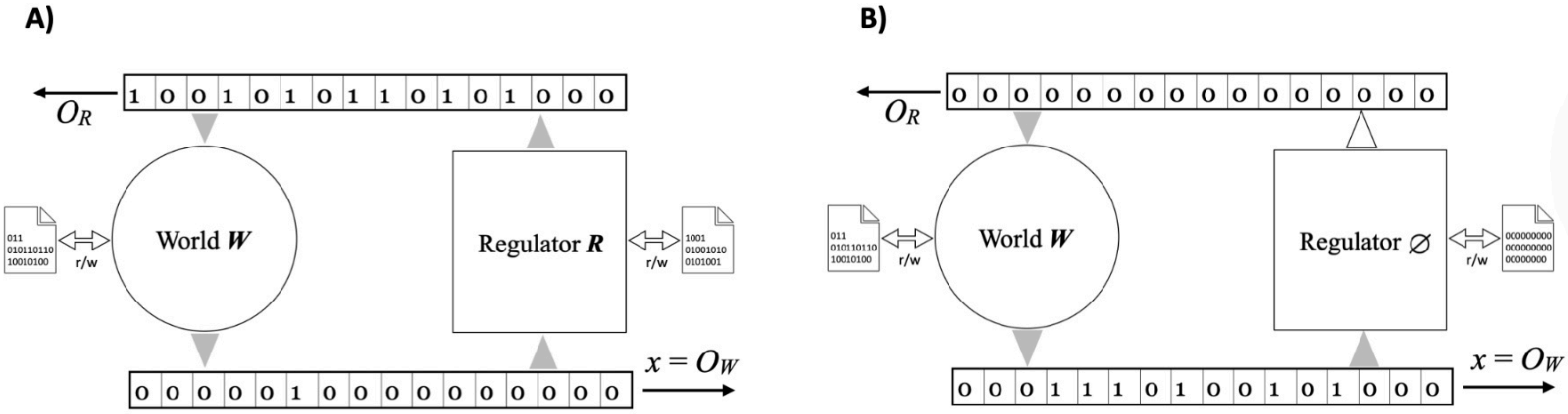
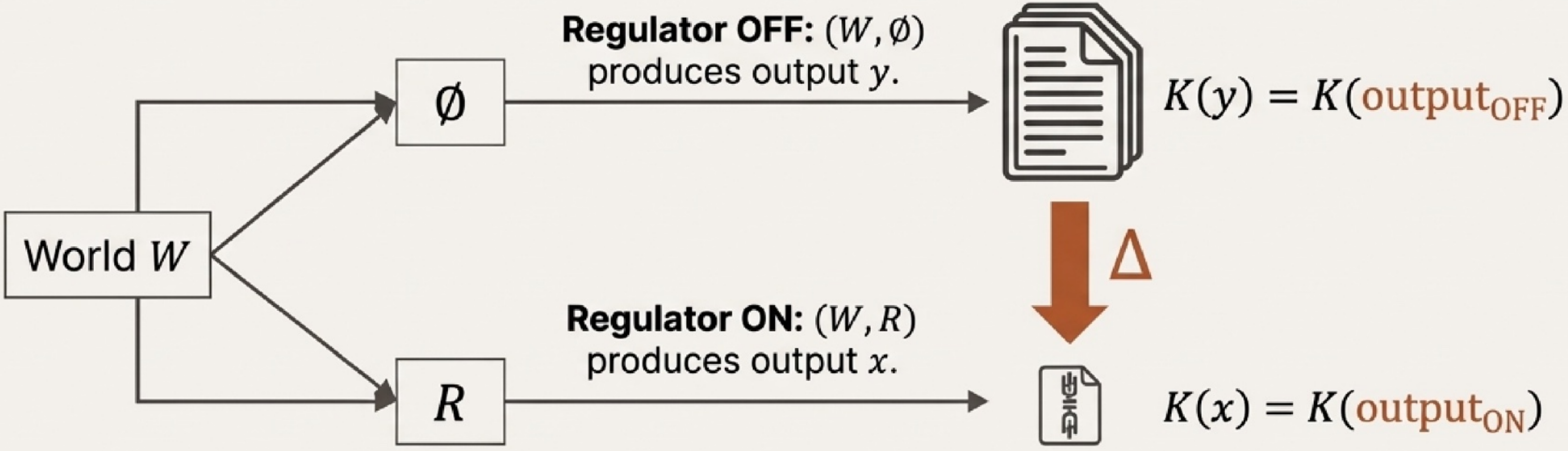


Figure 1: Regulation scenario. A) A good regulator R interacts with the world W so that the readout $x = O_W$ of the world's output is clamped to a simple, highly compressible sequence (e.g., almost all zeros). B) When the regulator is turned off, the output is more complex.

A Simple Test to Measure Success

We measure the regulator's effect by comparing two scenarios for a given World W.



The Compressibility Gap Δ

$$\Delta = K(\text{output}_{\text{OFF}}) - K(\text{output}_{\text{ON}})$$

Δ is the number of bits of compression achieved by the regulator. A "good" algorithmic regulator is simply one where $\Delta > 0$. The larger the Δ, the better the regulator.

The Algorithmic Regulator Theorem

Given that we observe a successful regulation (a large Δ), what can we infer about the relationship between the World W and the Regulator R?

$$P((W,R)|x) \leq C \cdot 2^{M(W:R)} \cdot 2^{-\Delta}$$

The Cost of Success

For every bit of compression you achieve (Δ), your explanation of that success becomes exponentially less likely by default. Seeing a simple output is surprising and requires a good explanation.

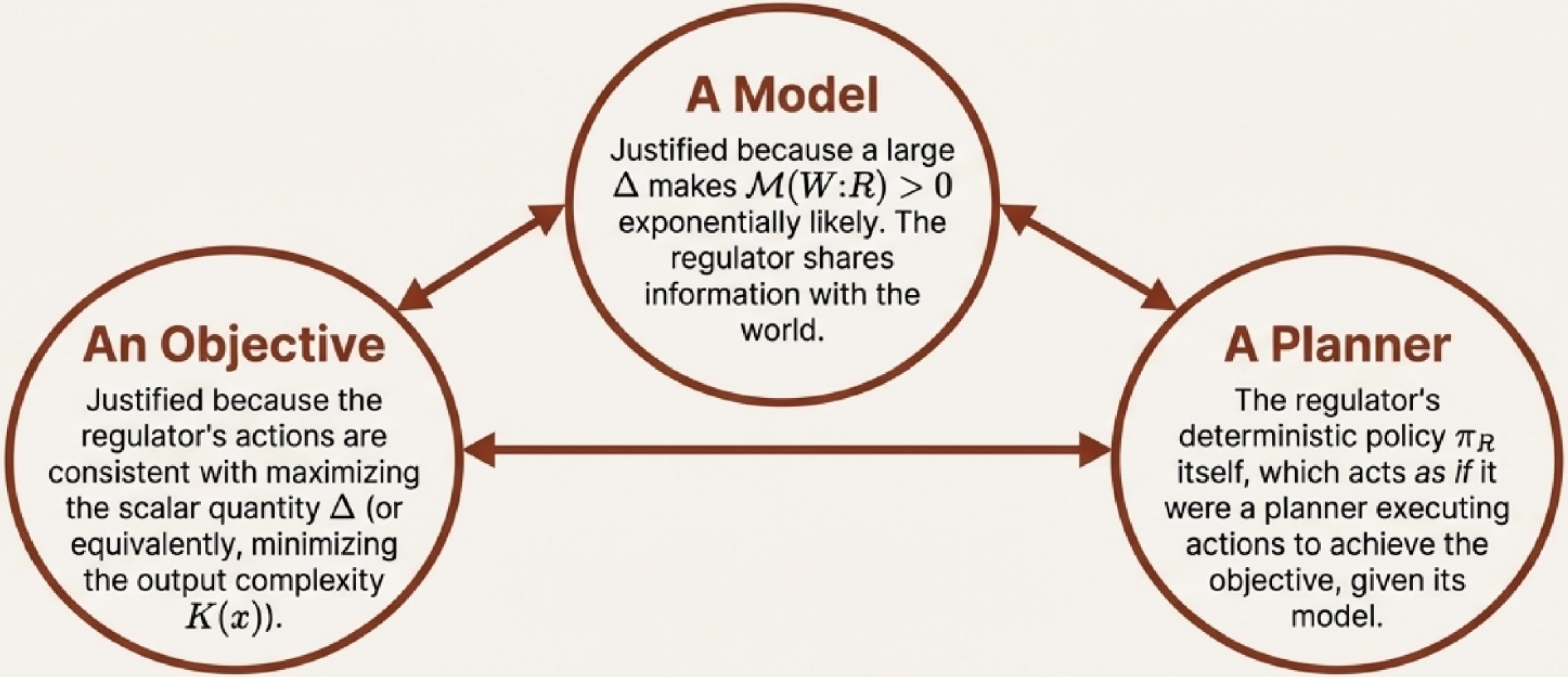
The Currency of a Model

The only way to overcome this exponential unlikelihood is for the World and Regulator to share information ($M(W{:}R)$). This term pays for the "cost of success."

Sustained, successful regulation (a large $\Delta > 0$) makes it exponentially unlikely that the regulator *doesn't* contain a model of the world.

From Evidence to Agency

The theorem provides grounds to infer an agent-like structure purely from observing a system's ability to compress its output. We can say the regulator behaves **as if** it were an agent with these components:



An 'As-If' Agent, Justified by Data.

A Complement, Not a Replacement

Aspect	GRT (Conant–Ashby)	IMP (Francis–Wonham)	A-GRT (This work)
Setting / Objects	Shannon entropy	LTI plants	deterministic prefix programs.
Definition of "model"	homomorphism	dynamical copy of exosystem	$M(W:\mathbb{R}) > 0$
Notion of "goodness"	minimize $H(Z)$	perfect asymptotic tracking	compressibility gap $\Delta > 0$
Core Theorem Statement	Every good regulator must contain a model of the system.	The internal model principle holds that regulation is possible if the regulator incorporates a model of the exosystem.	A regulator with a large Δ is exponentially unlikely to not contain a model of the world.
Scope / Use	Conceptual cybernetics link	Design backbone for robust regulation	distribution-free, single-episode diagnostics and a universal Occam calculus.
Information Source	Ensemble statistics	System dynamics	Observed data trace x
Mathematical Framework	Probability theory	Control theory	Algorithmic Information Theory

From Theory to a Testable Claim

The Challenge

Kolmogorov Complexity `K(x)` is a theoretical limit and is not computable.

The Practical Solution

We can use real-world, off-the-shelf compressors to get a practical upper bound on `K(x)`. We simply replace `K(x)` with `LC(x)`, the compressed length of `x` using a fixed compressor `C`.

The Experimental Recipe

1. Fix a lossless compressor `C`.
2. Quantize the system's readout if necessary.
3. Compute two code lengths: `LC(output_ON)` and `LC(output_OFF)`.
4. The difference, $\hat{\Delta}$ = `LC(output_OFF) - LC(output_ON)`, is your evidence. Persistent $\hat{\Delta} > 0$ is cumulative evidence of model content.

Examples of Practical Compressors

Classic

Lempel-Ziv family (gzip, lz4)

Algorithmic

Block Decomposition Method (BDM)

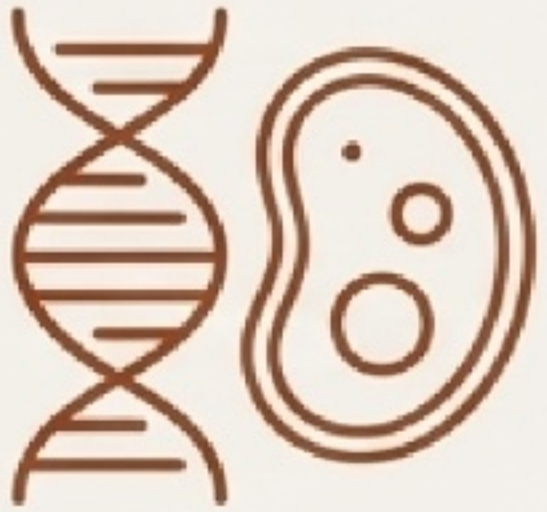
Modern

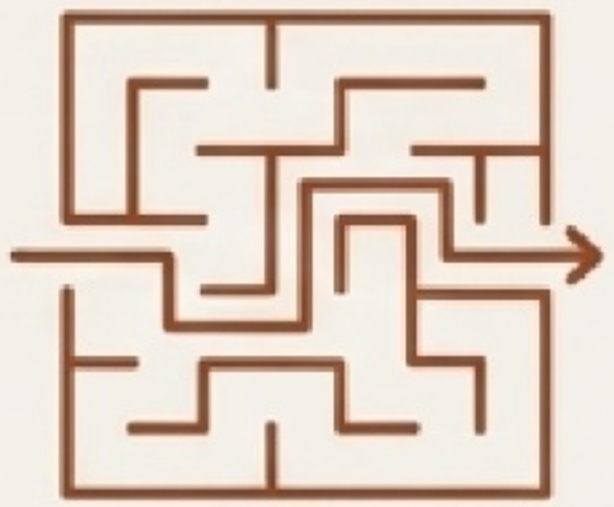
Learned compressors using Neural Networks (VAEs, Transformers)

A Universal Tool for Probing Agency

- **Distribution-Free**: It makes no assumptions about the underlying probability distributions of the world or signals. It works on single, individual sequences.
- **Architecture-Agnostic**: It does not assume linearity, an E/P split, or a particular causal structure. It only requires a computable $(W, \mathbb{R}) \to x$ mapping.
- **Universal**: It provides a principled, quantitative method to test for modeling and agency in *any* system where you can perform an ON/OFF experiment.

Domains of Application



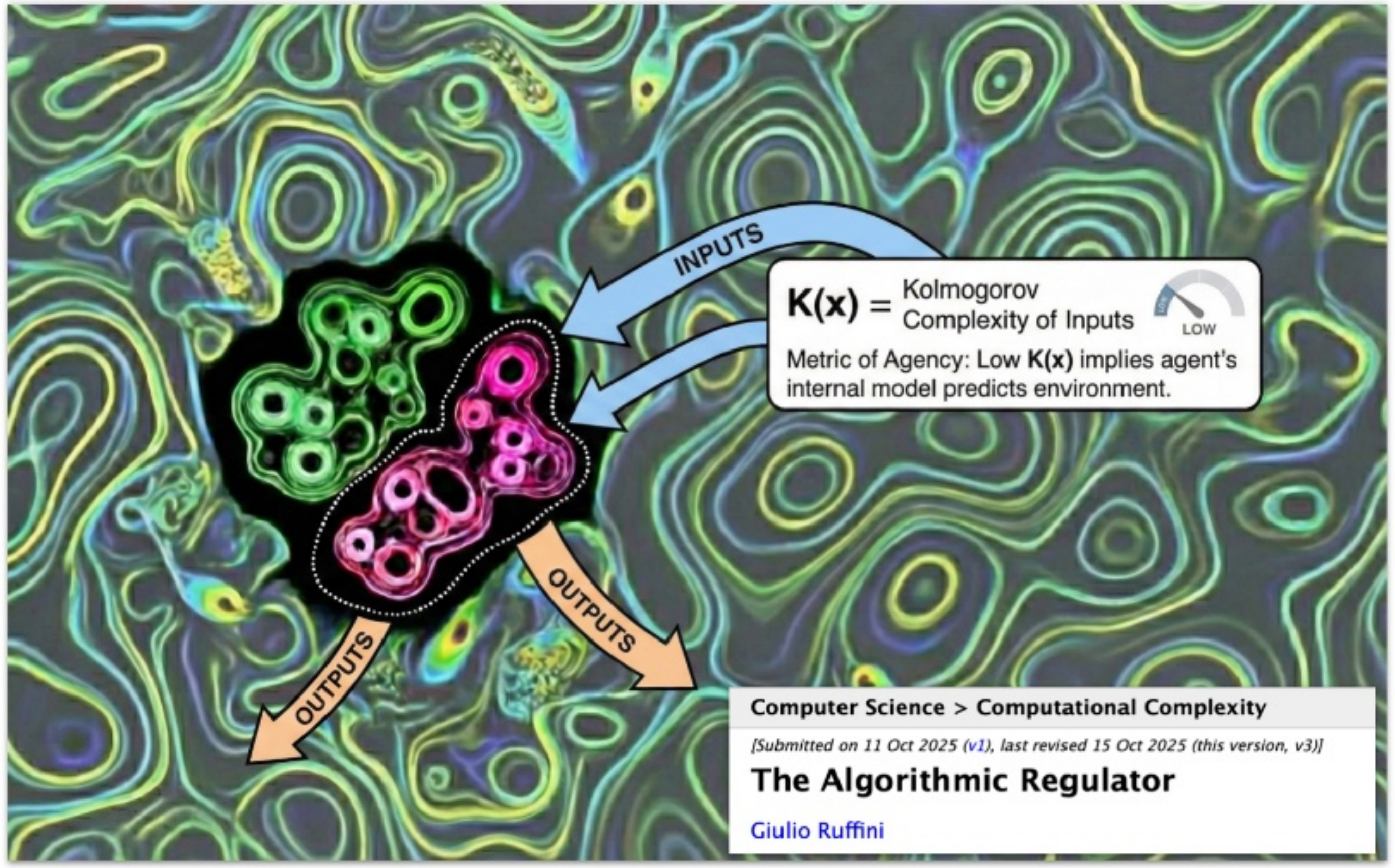
Simplicity is the signature of a model at work.

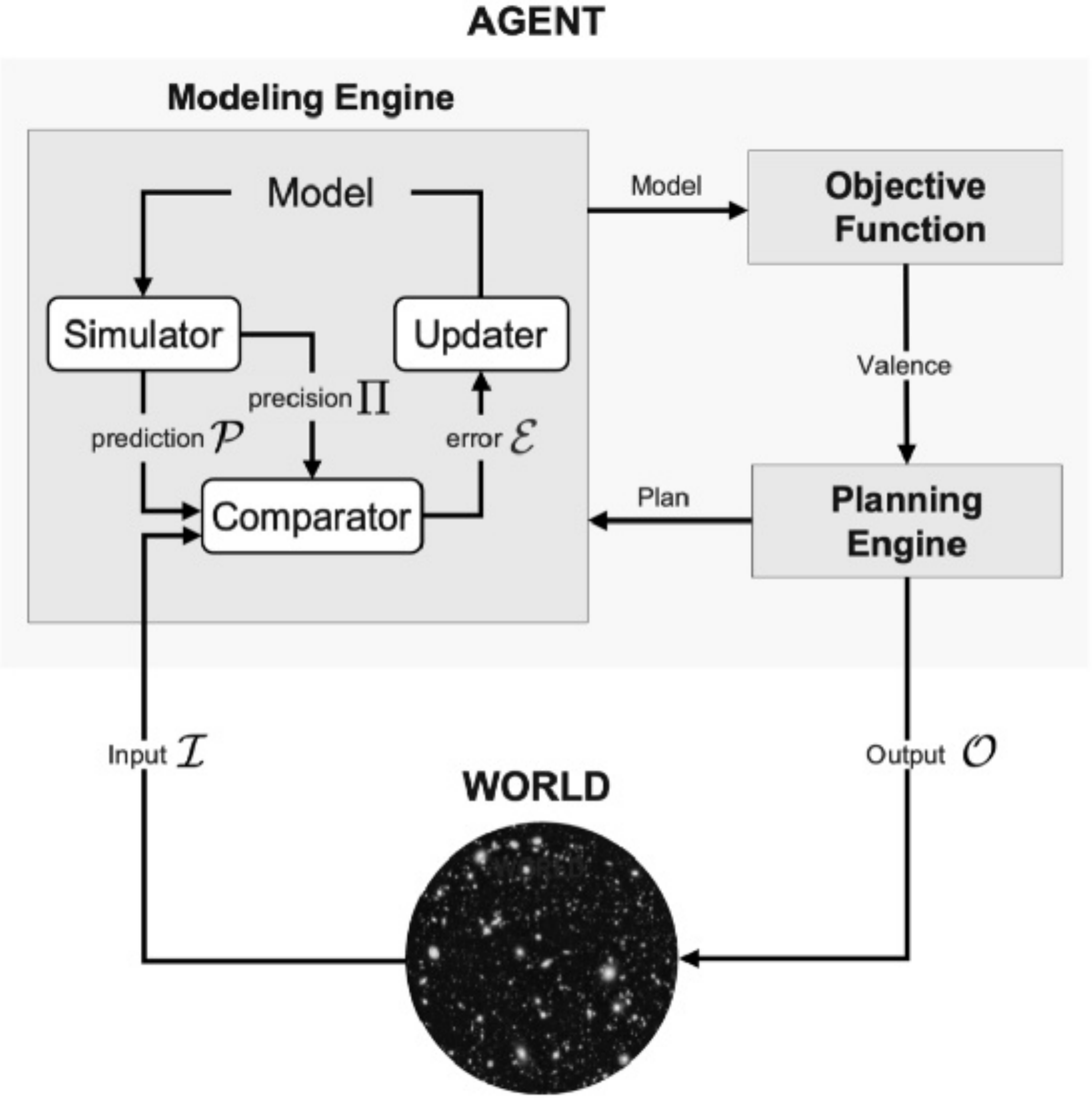
The more a system simplifies its world,
the more of that world it must contain.

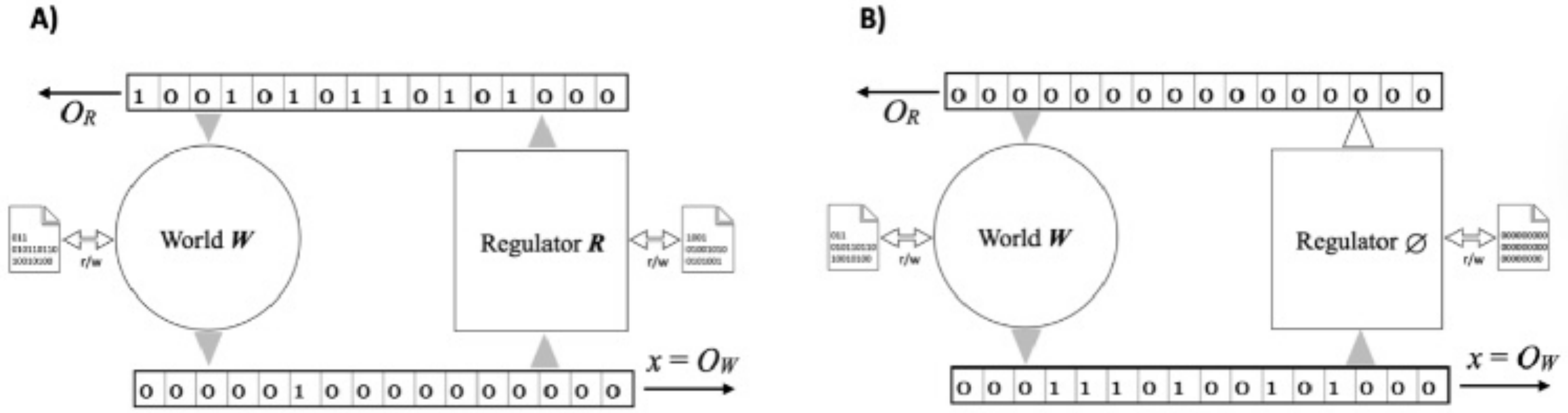
BCOM WP0021

Algorithmic Signatures of Agenthood in Artificial Life:
A Test of the Algorithmic Regulator in Lenia and Flow–Lenia
(draft)

Life and the Algorithmic Regulator (Ruffini 2025, arXiv)







Theorem 3.2 (Probabilistic regulator theorem). *Let $O^{(N)}_{W,R}$ and E^R_b be observed and let $\Delta := K(O^{(N)}_{W,\varnothing}) - K(O^{(N)}_{W,R})$. Then there exists $C > 0$ such that*

$$P\big((W,R) \mid O^{(N)}_{W,R}, \mathsf{E}^R_b\big) \;\le\; C \cdot 2^{M(W:R)}\, 2^{-\Delta}.$$

Equivalently, every bit by which $M(W{:}R)$ falls short of Δ costs a factor $\approx 2^{-1}$ in posterior support.

Every good regulator of a system must be a model of that system †

ROGER C. CONANT & W. ROSS ASHBY

Pages 89-97 | Received 03 Jun 1970, Published online: 08 Mar 2007

Computer Science > Artificial Intelligence

[Submitted on 2 Jun 2025 (v1), last revised 20 Oct 2025 (this version, v5)]

General agents contain world models

Jonathan Richens, David Abel, Alexis Bellot, Tom Everitt

Computer Science > Computational Complexity

[Submitted on 11 Oct 2025 (v1), last revised 15 Oct 2025 (this version, v3)]

The Algorithmic Regulator

Giulio Ruffini