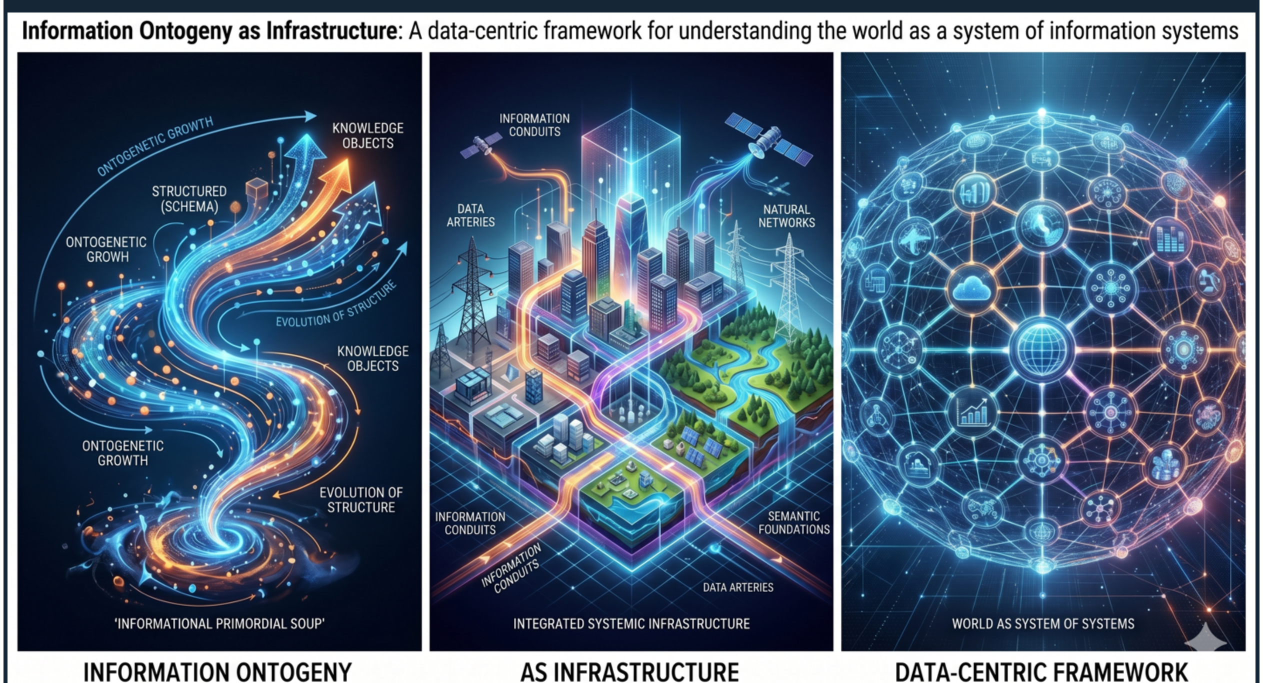
Information Ontogeny as Infrastructure

A Data-Centric Framework for Understanding the World as a System of Information Systems

Michael Ferrari, PhD
Founder & Managing Partner, Atlas Research Innovations
VP of Research, Moby · Senior Partner, AlphaGeo



Working Paper Version 4.0 · April 2026 · Complexity & Geospatial Intelligence Series

Author's Note: Two Traditions, One Claim

In late 2024, Dr. Fei-Fei Li, the computer scientist whose ImageNet dataset catalyzed the deep learning revolution, founded World Labs, a company built on a single proposition: that reasoning in AI systems is fundamentally constrained by what those systems can perceive. "For humans," Dr. Li wrote in her manifesto, "spatial intelligence is the scaffolding upon which our cognition is built. It drives our reasoning and planning, even on the most abstract topics." The company, which raised over a billion dollars within its first year, is engineering the consequence: AI systems whose outputs are not tokens or images but structured, physics-aware, generative representations of the world itself. These are 'World Models' whose capacity to reason is grounded in their capacity to perceive spatial and physical structure.

While Li arrived at her position from computer vision and AI, she also grounds her ideas in evolutionary biology and optics. I have arrived at the same position from spatial science, EvoBio, and complexity theory. The shared claim is: you cannot reason correctly about a system whose generative context you cannot perceive. Strip the perceptual ground from a reasoning system, and you have not simply reduced its precision; you have fundamentally changed the kind of understanding it is capable of producing.

This working paper develops that claim at a different scale and in a different register. Where Li is concerned with AI systems perceiving 3D physical environments, this paper is concerned with analytical systems understanding complex social-ecological-economic information environments. The technical apparatus is different, but the epistemological premise is identical. And the intellectual lineage behind this paper's version of the claim predates both traditions: it goes back to the biologist Susan Oyama, who argued in 1985 that information does not reside in a medium waiting to be read out. It develops through the interaction of an organism and its environment, and its character is constituted by that developmental history.

Oyama was writing about genes and embryos. Li is building machines that can see. This paper applies the same argument to the data systems that underpin our understanding of the world. The result is a framework called ***Information Ontogeny as Infrastructure***. Here I propose that the developmental history of information (*more particularly spatial information)* - how it came into being, through what relational context, across what spatial and temporal scales - is not incidental metadata. It is the primary substance of what the information ultimately represents and therefore should be viewed as the foundational lens through which complex systems can be understood.

Abstract

Standard data practice treats information as a static resource: collected, cleaned, modeled, consumed. In one sense, there is nothing fundamentally wrong with this assumption. Data is the raw material upon which all decisions are based. However, in this paper, I argue that treatment is constitutively misaligned to the better understanding of complex systems. In this view, information is not a resource, it has an ontogeny. Taking the analogy from biology, information then contains a developmental 'life-cycle' and history produced by the interaction of a measuring process and its environment. Further, this history is instructive with respect to what the information means. Data stripped of its developmental and organizational context is therefore less enlightening. It becomes a number rather than a signal as part of a structural relationship.

The central claim then is that information ontogeny functions as infrastructure: it is the foundational substrate that determines what relationships are knowable from a given data asset, what predictions are structurally possible, and what understanding is achievable. The world is best understood as a system of interacting information systems, each with its own ontogenic dynamics, coupled to each other through structural pathways that can be identified, characterized, and monitored. The data-centric spatial lens is the primary instrument for reading these pathways, because spatial provenance is the richest single key to an information asset's developmental history.

The epistemological position that reasoning is a function of perceptual grounding, and that grounding is constituted by ontogenic context is now being deployed at massive commercial scale in AI by researchers including Fei-Fei Li, whose World Labs is engineering AI systems whose reasoning is grounded in spatial and physical perception. From that tradition and from the complexity science tradition developed here, the same conclusion emerges: stripping the generative context from information does not simply degrade analytical performance. It makes a class of structural understanding structurally impossible. This paper attempts to describe what that class of understanding looks like, why it matters, and what it elucidates that other approaches can not find.

1. Information Has a Life

"Developmental information does not arise from nothingness. It develops always from the conditional transformation of prior structure - by ontogenetic processes."

- Susan Oyama, The Ontogeny of Information (1985)

In 1985, Susan Oyama published a slender and radical book in developmental biology. The prevailing dispute she addressed asked whether the information guiding an organism's development resided in its genes or in its environment. Oyama argued that the question was not framed properly at its foundation. Information, she insisted, does not reside anywhere. It is produced through the interaction of organism and environment. It has a developmental history, or an ontogeny, that is not separable from what the information is and what it does. The gene is not the carrier of pre-formed biological meaning; it is one participant in a developmental process through which meaning is produced.

The extension of this insight to data is direct and overdue. Information is not a carrier of pre-formed meaning waiting to be read out of a database. It is the product of a generative process which is expressed as the interaction of a measuring instrument with a physical or social phenomenon situated in a specific context. That context is not a metadata label appended to the observation. It is actually the condition of the origination of the data, and therefore part of what the information is and what it represents. If you strip that context, you have not simplified the data, you have amputated the part of it that encodes structural meaning.

Here we can apply the four stages of biological ontogeny to spatial data.

1.1 Birth: The Generative Context

Every piece of spatial data is produced by a process located in the world. A satellite sensor reading, a shipping manifest, a clinical measurement, a weather station observation, a commodity futures price, a regulatory filing…each of these emerges from the interaction of a measuring instrument with a phenomenon that is situated in space, time, institution, and ecology. That situation is the datum's birthplace, and the birthplace encodes a dense bundle of structural context.

A nitrogen dioxide reading from Sentinel-5P over the Yangtze River corridor encodes Chinese manufacturing activity, regional atmospheric dynamics, the regulatory environment governing industrial emissions in that province, the seasonal climate system, and the coupling of that location to global atmospheric teleconnection pathways. None of this is written in the numerical value of the reading itself. All of it is encoded in the reading's spatial and temporal provenance. Preserve that provenance and the reading is a signal indexing a complex of structural relationships. Strip it (as we are seeing with the current dismantling of scientific observation information collection) and the reading is simply reduced to a number that can be correlated with other numbers, but cannot be placed in the world.

The birthplace is primary. It is the first and most generative key to an information asset's meaning, and its recovery or preservation is the first task of any analysis that aspires to structural understanding.

1.2 Development: Relational Meaning Through Contact

Information does not arrive at its full meaning at the time of it's birth. It develops meaning through contact with other information systems. A monsoon anomaly in the Bay of Bengal acquires pharmaceutical significance through its structural coupling to water availability at manufacturing facilities in Hyderabad. A crop stress index in Brazil acquires meaning for antibiotic production through the fermentation feedstock dependencies of pharmaceutical manufacturing. A port congestion signal in the South China Sea acquires meaning for retail inventory dynamics through the supply chain networks that route through it.

None of these relational meanings is intrinsic to the originating signal. Each is produced by the structural coupling between information systems. This is the developmental contact between a signal and the network of other signals and processes it encounters as it moves through the world. The same data, moving through and across different relational networks, acquires different meanings. So the context is the mechanism of meaning production, and it operates throughout the information's developmental trajectory, not only at its birth.

1.3 Maturity: The Actionability Window

Information, like biological fitness, experiences a maturity window. This is the period of peak predictive value during which its relational meaning is sufficiently developed to be actionable. Before that window, the signal may be detectable but lacks the relational context to be interpretable and actionable. After that window, the information ages out of relevance as the process it indexed evolves beyond the snapshot it captured.

This maturity window is itself also structurally determined. A sea surface temperature anomaly in the tropical Pacific reaches peak predictive value for downstream agricultural outcomes through a propagation sequence (***atmospheric teleconnection dynamics***) that unfolds over weeks to months, and in some cases, years. That propagation sequence is the information's developmental trajectory; the maturity window is its peak; the characteristic lead time between the originating signal and receiving-system outcome is the expression of its developmental velocity. Understanding information maturity means understanding the developmental dynamics of the system that produced it, not simply estimating a statistical lag.

1.4 Senescence: Preparing for the Next Generation

Senescence is the final stage in the information lifecycle, which allows data to deconstruct and 'fade' and in the process provide the raw material to spawn the next generation of information. Information ages at rates set by its generative system's dynamics: financial market data in milliseconds, satellite cropland imagery across growing seasons, infrastructure survey data across years of physical change. Old information also reactivates: a 150-year sea surface temperature time series is the developmental record of a climate system whose long-period variability may now the most important predictive feature for infrastructure planning over decadal timescales.

> *Information ontogeny is a description of how data actually acquires, sustains, and loses meaning in complex systems. Treating data as a static resource rather than a developmental trajectory is the primary source of structural blindness in contemporary analytical practice.*

2. Ontogeny as Infrastructure

"There appears to be a growing decoupling between the conditions that infrastructures were designed for and today's rapidly changing environments."

- Mikhail Chester and Braden Allenby

Infrastructure is a particular kind of foundational system: one whose value lies not in what it does directly but in what it enables other systems to do. A road network does not move goods; it enables goods to be moved. An electrical grid does not compute; *it enables computation*. Infrastructure shapes what flows, at what speed, with what reliability, and with what vulnerability to disruption. Its defining characteristic is that it is generative: it determines what is structurally possible and what is structurally impossible for the systems it serves.

Information ontogeny functions as infrastructure in exactly this sense. The developmental history of information - *the chain of generative contexts, relational contacts, maturity trajectories, and decay rates that constitute its meaning* - this serves as the substrate that determines what is analytically and physically possible. When that substrate is preserved and analyzed, structural relationships become visible that would otherwise be inaccessible. Again, when it is stripped, as it is by most current data pipelines in operation, the analytical capacity of the resulting system is degraded in kind, and an entire category of understanding becomes structurally impossible.

2.1 Ontogenic Infrastructure Shapes What Is Knowable

Even as an engineering construct, physical infrastructure is not neutral. A road network does not simply connect existing destinations; it determines where destinations form, where economic activity concentrates, and where isolation persists. It is an enabler of resource transfer, as well as disease transmission. The configuration of infrastructure determines what is accessible and what is not, and therefore what is possible and what is not. Ontogenic information infrastructure operates the same way, or at least it should.

The degree to which an information asset's developmental history is preserved determines what structural relationships are visible and what are inaccessible. An organization that retains the spatial provenance of its transactional data, the institutional context of its regulatory filings, and the temporal provenance of its sensor readings has access to a class of structural inference that an organization that strips these attributes does not. This difference is not quantitative. It is the difference between being able to see a category of relationship and being constitutively blind to it. Ontogenic infrastructure defines the epistemic terrain of what can be known.

2.2 The Dark Data Problem, Reframed

A great deal of data that is collected is never used. This is typically interpreted as a data management problem: the data exists but is difficult to find and access. Building over a layer with 'smart' ontogenic perspective reframes it more precisely and more consequentially.

When surveying the commercial data landscape, a pattern is evident that is consistently repeated, regardless of organization's size and data-sophistication: most enterprise data is ontogenically dark. It has been collected in ways that stripped its developmental context. The spatial provenance was never recorded, it has been eliminated, or has simply been lost. The temporal sequence has been compressed into aggregates that erase the dynamics that were originally indexed. As a result, the institutional and relational context of production has not been retained. So the data then exists simply as numbers, but the information - *the structured meaning that can be related to other systems through its developmental history* - was never fully exploited. If we refer to the data lifecycle described above, these data assets are collected but not born, in the ontogenic sense.

It should then be no surprise that dark data recovery is therefore not a retrieval problem but an ontogenic reconstruction problem. Some of what was lost can be recovered, think spatial attributes inferred from partial records, temporal sequences partially reconstructed from audit trails and/or statistical interpolation, but the recovery is always incomplete. The argument for building ontogenic infrastructure at the point of data collection, rather than attempting reconstruction after the fact, is the same argument for laying pavement for roads before a city is built, not after.

2.3 Space as the Primary Developmental Dimension

Enter Space.

Of all the dimensions of information's generative context - spatial, temporal, institutional, relational - space is emerging as being analytically paramount. This is not because place is more important than time or institutional context in any absolute sense. It is because a location in space encodes, in a single coordinate pair, the densest available concentration of generative context.

We can illustrate this through climate. A spatial coordinate encodes variables that constitute the climate system in space and time as the information is subject to teleconnective couplings to other biological, physical, chemical and economic systems; the ecological regime in which it sits, the regulatory jurisdiction it falls within, the physical infrastructure it has access to, the economic geography of markets and supply chains that shape its context, and the institutional history of the specific place. No other single attribute encodes a comparable density of generative context in a bias-free manner. This is why the data-centric spatial lens is the appropriate instrument for reading information ontogeny: spatial provenance, derived from space, is perhaps the richest single key to unlocking an information asset's developmental history.

> *The data-centric spatial lens is not a methodology for making maps. It is a methodology for reading the developmental history of information, using spatial provenance as the primary key. The map is the output. The understanding of relational structure is the goal.*

3. Information Systems of Systems

"The universe is made of energy, matter, and information, but information is what makes the universe interesting."

\- Cesar Hidalgo, Why Information Grows

The world is not a single complex system, it is a system of systems. A configuration of partially overlapping, structurally coupled, independently evolving information systems, each with its own spatiotemporal dynamics, each generating information that flows into and shapes the others. Understanding the world analytically means understanding this configuration: which systems are coupled to others, through what directional and structural transfer pathways, at what timescales, and with what developmental consequences for the information that traverses those routes.

This is a stronger claim than the standard assertion that 'everything is connected." Connectivity is a structural fact. What the information *systems of systems framework* adds is an account of the informational dynamics of connection: how information born in one system develops relational meaning as it traverses coupling pathways, how it matures into actionability at specific points in its journey, how it arrives transformed in receiving systems and influences their behavior. These developmental dynamics are the substance of what connections actually do, they and supersede questions of 'what' and move towards questions of 'where and why'.

3.1 Generative Systems and Ontogenic Transfer Pathways

Information systems, in this framework, are not simple software platforms or databases. They are generative processes that produce information with characteristic ontogenic properties: specific spatial and temporal scales of production, specific relational coupling pathways to other systems, characteristic maturity windows and decay rates.

Re-emphasizing the climate system - the global monsoon system is a physical information transfer system: it produces temperature, precipitation, and atmospheric dynamics signals at characteristic spatial scales (subcontinental to continental) and temporal scales (seasonal to interannual), with well-documented coupling pathways to other climate systems and to agricultural, hydrological, and infrastructure systems. The global pharmaceutical trade network is a different type of commercial information system: it produces shipping manifests, production records, regulatory filings, and inventory dynamics that index the state of global drug supply at characteristic spatial scales (facility to regional to global) and temporal scales (daily to quarterly). A sovereign bond market is an information system. The emerging network of IoT sensors embedded in physical infrastructure is an example of a physical-technological-human information system.

Coupling between information systems operates through ontogenic transfer pathways: structural channels through which information born in one system develops new meaning as it enters another. These pathways are determined by physical geography (shared climate zones, shared transportation corridors, shared water systems), institutional architecture (regulatory jurisdictions, trade agreements, legal frameworks), ecological structure (shared biological production chains, shared environmental dependencies), and the physical-digital infrastructure that increasingly mediates all of these. ***Understanding the pathways is understanding the developmental mechanisms of the inter-system couplings***.

3.2 Scale, Nesting, and the Granularity of 'Meaning'

An information systems of systems framework needs to embed flexibility with respect to the fluid characteristics of scale and nesting.

A crop stress signal from a MODIS satellite pass can be read at the level of a specific field, a county, a watershed, a national agricultural system, or a continental production zone. Each level of spatial aggregation produces a different piece of information with a different ontogenic character: different relational couplings to other systems, different maturity windows, different decay rates, etc. A field-level signal may be most actionable for a farm operator on a daily to one-week timescale. The continental-level aggregate is most actionable for a commodity trader on a three to six-month timescale. The national-level estimate is most actionable for a pharmaceutical procurement officer planning fermentation input sourcing across a growing season.

However, these are not independent pieces of data. They are nested stages in a single ontogenic hierarchy: the field-level signal is developmentally prior to the county-level aggregate, which is prior to the watershed-level integration, which is prior to the national production estimate. Meaning is produced at each transition, and meaning is also *irreversibly lost* through aggregation, as this is a form of ontogenic compression that discards some of the developmental history of the original signal. Economically, we know why compression and aggregation exist - but practically more thought needs to be applied to 'why'. The question of what scale to analyze information should therefore be determined by the scale of the decision it is intended to inform, and the ontogenic pathway between originating signal and analytical output should be explicitly traced and preserved. Same as physical infrastructure.

3.3 Boundaries as the Sites of Maximum Ontogenic Activity

In a system of information systems, **the boundary zones,** where one information system transitions to another, are the most analytically productive territory. These are the sites where ontogenic information transfer occurs; ie., where information born in one generative context undergoes transformation as it enters a new relational environment. They are also the sites where new meaning and context layers are produced, where coupling mechanisms are most legible, and where the structural relationships between distant systems first become visible.

The boundary between the climate information system and the agricultural information system is where atmospheric dynamics are translated into soil moisture, heat stress, and growing-season anomalies. The boundary between the agricultural information system and the pharmaceutical information system is where crop production anomalies translate into fermentation feedstock availability and cost, which translate into antibiotic and vitamin production constraints. The boundary between the pharmaceutical information system and healthcare delivery systems is where drug supply dynamics translate into patient access outcomes. Each boundary is a site of ontogenic development: information arrives from one system, undergoes developmental transformation through the coupling mechanism, and departs as information of a different character, carrying new relational meaning, into the receiving system.

The full developmental chain from an atmospheric signal to, for example, a healthcare outcome crosses at least four information system boundaries, each mediated by a specific coupling mechanism. No single-domain analytical framework can trace this chain. Understanding it requires a framework that treats the chain itself, which is the sequence of boundaries and the ontogenic transformations at each, as the primary object of analysis.

3.4 Teleconnections as Developmental Signatures

Teleconnections are the observable signatures of coupling between distant information systems. In climate science, an El Niño event in the tropical Pacific is coupled to monsoon anomalies in South Asia, drought in Australia, and winter temperature anomalies in North America. The teleconnection is not a statistical correlation between variables in these distant regions. It is the developmental trace of information, essentially physical energy transfer, born in the tropical Pacific ocean-atmosphere system as it traverses the global atmospheric information system, through the coupling mechanism of Rossby Wave propagation, and arrives, transformed but structurally related, at receiving systems on multiple continents.

This developmental framing is essential for understanding why well-characterized teleconnections have durable predictive power while statistical correlations often do not. A correlation is identified in historical data and applied as a predictor. It persists as long as the underlying coupling remains stable and it fails when the system reorganizes. An understanding of shifting baselines, and how to interpret them, is key. A teleconnection grounded in an identified coupling mechanism where the ontogenic transfer pathway is characterized and not merely the correlation between endpoints is robust to many forms of system reorganization. This is because the mechanism, not the historical correlation, is the source of its predictive content. When the system reorganizes, the mechanism changes in ways that are observable and recharacterizable. While the framework updates, the correlation simply fails.

Moving to a practical example, supply chain teleconnections operate by exactly this logic. The sea surface temperature anomaly in the Bay of Bengal is coupled to Indian monsoon dynamics, which are coupled to water availability at pharmaceutical manufacturing clusters in Andhra Pradesh and Telangana. These are coupled to active ingredient production volumes for the global generics market. Each coupling is an ontogenic transfer pathway. The signal travels through a developmental chain, acquiring new meaning at each stage. The predictive value of the originating signal for the terminal outcome, in this case a drug supply disruption in distant markets, derives from the structural integrity of those pathways, not from the historical statistical association of the variables at the chain's two ends.

3.5 The Physical-Digital Layer and Accelerated Ontogeny

The emergence of a globally distributed layer of digital infrastructure has not simply added a new information system to the world. It has created a new class of ontogenic transfer pathway that operates at timescales orders of magnitude faster than the physical coupling pathways it overlays, and in doing so has fundamentally altered the developmental dynamics of every information system it touches.

Before this digital layer, information born in a physical system propagated through physical coupling pathways: commodity trade flows, population movement, infrastructure dependencies. These pathways had characteristic timescales of weeks to months for commodity markets, months to years for population adjustments, years to decades for infrastructure transitions. The ontogenic transfer was slow enough that developmental trajectories were often visible and the information's provenance could be tracked.

Recently, the digital layer has collapsed these timescales, but has done so non-uniformly. A satellite observation of a flood in a major agricultural region propagates into commodity futures markets within minutes through algorithmic trading systems. The ontogenic transfer, from physical event to market price signal, has now been compressed from days to minutes.

But crucially, the relational meaning of the resulting price signal still carries the spatial birthplace of the original observation: a flood in a specific place, indexing specific crop dependencies, with specific downstream consequences. The developmental velocity has changed; the spatial structure of the meaning has not.

What is genuinely new is the emergence of information systems that are native to the digital layer: social media dynamics, algorithmic market behaviors, platform-mediated demand patterns, AI-generated signals. These systems produce information with ontogenic properties that have no physical analogue. The knowledge created is born via digital processes, develops through digital coupling, and decays at digital timescales. Knowledge transfer is now coupled to physical information systems in both directions, creating feedback dynamics that neither layer exhibits independently. The physical-digital boundary is the most rapidly evolving information system boundary currently active, and it is the site of some of the most consequential and least understood ontogenic processes in the contemporary world.

> *The digital layer does not represent the physical world. It participates in it. Ontogenic transfer pathways now operate simultaneously through physical and digital couplings, at different timescales, producing information whose meaning is constituted by both. An analysis that treats these layers as separable is not simply incomplete. It is misconfigured for the system it is trying to understand.*

4. The Perception-Reasoning Dependency

"Spatial intelligence is the frontier beyond language - the capability that links imagination, perception, and action."
— Fei-Fei Li, World Labs Manifesto (2025)

The argument developed in the preceding sections, that information ontogeny is foundational to meaning, and that stripping ontogenic context makes a class of structural understanding impossible, has an exact parallel in a very different tradition. Fei-Fei Li, whose ImageNet project gave the deep learning revolution its primary training substrate, has spent the past several years arguing that language-based AI systems are subject to a fundamental limitation: they cannot reason correctly about the world because they cannot perceive it. They process tokens that represent the world, but tokens stripped of the spatial, physical, and geometric context that gives worldly phenomena their structure. The consequence is not that these systems are slightly wrong. Instead, Dr. Li argues that they cannot achieve a category of understanding that requires perceptual grounding.

Li calls this category 'spatial intelligence' and she founded World Labs to engineer its computational realization: AI systems whose reasoning is grounded in structured, physics-aware, generative representations of 3D space; in other words, world models. The goal is not better image generation. The goal is reasoning that is grounded in the geometry and physics of the world, and therefore capable of a kind of inference about spatial and physical relationships that token-based systems cannot approach. This is how babies learn and reason about the world. And the logic makes sense.

Also, the structural parallel to the information ontogeny framework is precise, and it is worth stating explicitly.

4.1 The Isomorphism

Li's argument, translated into the language of this paper: a system whose training data has been stripped of spatial and physical context (ie., *reduced to tokens*) cannot reason correctly about the spatial and physical world. The stripping of perceptual context does not simply reduce precision. It makes a class of understanding structurally inaccessible.

This paper's argument: an analytical system whose data has been stripped of ontogenic context (*ie., reduced to numbers*) cannot reason correctly about the information systems of the world. The stripping of developmental context does not simply reduce precision. It makes a class of structural understanding structurally inaccessible.

I believe the isomorphism to be exact. In both cases, the failure mode is the same: a reasoning system operating on context-stripped representations cannot access the structural relationships that are only visible in the full developmental context of the information it is processing. In both cases, the corrective mechanism is the same: restore the perceptual spatial, physical and ontogenic grounding that was stripped in preprocessing, and the reasoning capacity recovers a category of understanding that the stripped version could not achieve.

The scale of application is admittedly different: Li is concerned with 3D physical environments at the scale of objects and rooms; this paper is concerned with complex social-ecological-economic information systems at the scale of supply chains, climate systems, and global markets. And the technical apparatus is also different: World Labs is building neural world models; this paper uses geospatial analysis, network science, and complexity theory. But the epistemological premise is shared and, I would argue, fundamental.

4.2 The Sim-to-Real Gap as an Ontogenic Failure

World Labs' work provides one of the most vivid and practically important illustrations of the ontogenic failure the paper diagnoses. The 'sim-to-real gap' is the name given to a pervasive problem in robotic AI: a system trained in simulation fails to generalize to physical reality, often catastrophically, due to slight variations in lighting, surface texture, object positioning, or environmental physics that the simulation did not capture. The roots of this problem can incidentally be traced to early climate modeling efforts via Ed Lorenz' initial observations of initialization error in complex dynamical systems yielding wildly different results. The simulation and the physical world produce data that looks the same in many surface respects but was born through radically different generative processes. The simulated world has a different ontogenic history than the physical world.

It follows that the model trained on simulated data has learned from ontogenically stripped information: data that was born in a procedurally generated environment that lacks the physical, material, and historical generative context of the real world. Synthetic data is great, until it isn't. When that model is deployed in the real world, it encounters information whose birthplace it has never seen and whose generative context it cannot read. The failure is not a failure of model capacity. It is a failure of ontogenic grounding. Until synthetic data can construct a synthetic ontogeny, this will remain a roadblock.

This is identical in structure to what happens when a supply chain model trained on historical data is deployed into conditions produced by a structural reorganization it has never encountered. The historical training data was born in a system configuration that no longer exists. The deployment data is born in a new configuration whose generative context the model has not been exposed to. The model fails not because it lacks parameters or equations, but because it cannot read the ontogenic context of the information it is now receiving. In both cases, the corrective is the same: build analytical systems that are grounded in the structural character of their information's generative context, not only in the statistical patterns of its historical outputs.

4.3 World Models as Formalized Ontogenic Substrates

The world model concept that Li and World Labs are developing is, in the language of this paper, an engineered formalization of information ontogenic infrastructure at the AI system level. A world model is a structured representation of the generative context of visual and physical information: it preserves the geometry, physics, and semantic structure of the environment that produced the training data, and it enables the model to reason about spatial and physical relationships rather than merely pattern-match on pixel or token sequences.

In the paper's terms, a world model is an attempt to preserve the ontogenic substrate of visual information so that the reasoning system built on it can access the structural relationships that are only visible when generative context is retained. This is exactly what the information ontogeny framework asks of analytical systems operating on social-ecological-economic data: not just that data be collected and processed, but that the developmental history of the data be preserved and made available to the analytical system as primary structure, not stripped out as preprocessing overhead.

The technical implementation is necessarily different across the two domains. World Labs is building neural architectures that learn 3D world representations from visual data. The information ontogeny framework is building analytical architectures that preserve and analyze the spatial provenance, temporal trajectory, and relational coupling history of complex systems data. But the functional goal is the same: an analytical system whose reasoning is grounded in the generative context of its information, and that therefore has access to a category of structural understanding that context-stripped systems cannot achieve.

5. What This Methodology Finds That Others Cannot

The information ontogeny framework I am describing leads to a specific category of finding that is structurally inaccessible to the two dominant alternatives: pure AI-driven pattern recognition and pure domain expertise. The reason for each inaccessibility is now clear from the preceding argument.

5.1 What Pure AI Cannot Find

Contemporary AI systems are powerful pattern recognizers operating on ontogenically stripped training data. Spatial provenance is typically reduced to numeric coordinates. Temporal sequence is reduced to lag features, and institutional context is intentionally absent. The developmental history of the data - which information systems it passed through, what relational transformations it underwent - is not preserved. As a result, the model learns from what the context-stripped data contains. It cannot find the structural relationships that require ontogenic context to be visible, because that context was removed before the model encountered the data.

The deeper failure is structural, not technical. AI systems trained on historical data reason about the world's past configuration. They cannot detect structural reorganizations before those reorganizations appear in the training record, because reorganizations are, by definition, departures from historical patterns. A model trained before 2020 had no basis for anticipating the pharmaceutical supply chain vulnerabilities exposed by COVID. This is not because the data was insufficient, but because the structural concentration of pharma manufacturing in China was a feature of the network's topology, not of its historical output statistics. The information ontogeny framework reads network topology as a primary feature; AI pattern recognition reads historical output statistics. These are not competing approaches to the same information. They are accessing different information entirely.

5.2 What Domain Expertise Cannot Scale

Deep domain expertise provides irreplaceable for inter-and-intra-system mechanism understanding. A specialist in country-specific pharmaceutical regulation understands the institutional dynamics of that system with depth and nuance that no algorithm approaches. But the same expert cannot, working within that single domain, systematically trace the ontogenic transfer pathway from Bay of Bengal sea surface temperatures through monsoon dynamics to Hyderabad water availability to global generics supply, to repeat the example discussed above. That cross-system, cross-scale developmental chain is not accessible through depth of expertise in any single information system. It requires the ability to trace the chain across boundaries, which requires a framework that treats information systems and their couplings as the primary object of analysis.

But domain expertise carries the limitation of its interpretive framework. The pharmaceutical supply chain expert reads data through pharmaceutical supply chain categories. The climate scientist reads data through atmospheric dynamics categories. Neither, working independently, has the framework to recognize that the same structural coupling mechanism that produces El Niño teleconnections in the climate system produces analogous teleconnections in supply chains and financial networks. The information ontogeny framework provides the cross-domain scaffolding: a common language of generative systems, ontogenic transfer pathways, and developmental timescales that makes systematic cross-system analysis tractable. Knowledge is created at the boundaries.

5.3 What the Ontogenic Lens Specifically Reveals

Three categories of finding are structurally accessible to the information ontogeny framework and inaccessible to the alternatives:

- **Self-organized structural concentrations of risk.** These hubs are visible from network topology before they manifest as disruption. The geographic concentration of supply chains and/or manufacturing is a self-organized feature of the supply chain upstream/downstream information network that is readable from its topology. This is reading the ontogenic history of the industrial information systems as they operate, not its historical outputs.
- **Teleconnective chains with mechanistic grounding.** These provide lead times that are structural rather than correlative. Again, the developmental chain from Bay of Bengal sea surface temperatures through Indian monsoon dynamics to end-user/manufacturer are structurally grounded ontogenic transfer pathways with a characteristic developmental timescale of 6 to 18 months depending on the stage. Lead time is not a statistical lag estimated from historical data. It is the developmental velocity of a specific coupling mechanism, and it is therefore durable across system reorganizations in ways that pure correlations are not. An organization that instruments this chain, for example companies monitoring the signal at its source and understanding the developmental pathway through which it reaches its decision-relevant stage, has a structural early-warning capability that no amount of historical data analysis provides.
- **Regime change signals before disruption.** Visible in spatial pattern changes that precede shifts in system output statistics. Complex systems reorganize before they produce anomalous outputs. The reorganization is detectable in changes to the spatial pattern of information flows: shifts in the geographic distribution of manufacturing activity, changes in trade network topology, anomalies in the spatial correlation structure of climate and agricultural signals that precede regime transitions. These are signals in the ontogenic substrate of the information systems. Changes in how information is being produced, not yet in what it is saying. Reading these signals requires an analytical framework that treats the developmental structure of information as primary.

6. Conclusion: How to Learn Spatially

"The most important question is not 'what does the data say?' It is 'where was this data born, and what has happened to it since?'"

— Michael Ferrari

The argument of this paper is a sequence of nested claims, each following from the one before. Information has a developmental history - *an ontogeny* - that is constitutive of its meaning, not separable from it. That ontogeny functions as infrastructure: it determines what is structurally knowable from a given data asset. The world is a system of interacting information systems, each with its own ontogenic dynamics, coupled through transfer pathways that are spatially organized and structurally characterizable. Space is emerging as the primary key to reading an information asset's developmental history, because a spatial coordinate encodes the densest available concentration of generative context. A methodology that preserves and analyzes information ontogeny surfaces structural relationships, which include long-range physical teleconnections with durable predictive power, that neither AI pattern recognition nor domain expertise can find, because both operate on ontogenically stripped representations.

This framework has an independent validation from a tradition that reached a similar conclusion from a completely different starting point. Dr. Fei-Fei Li's work at World Labs instantiates the perception-reasoning dependency at the AI engineering level: reasoning, Dr. Li argues, is a function of perceptual grounding, and perceptual grounding requires the structural context of the physical world to be preserved in the representations the reasoning system operates on. The sim-to-real gap is the engineering expression of the same failure mode this paper diagnoses theoretically: a system trained on ontogenically stripped information cannot reason correctly about a world whose information carries a different developmental history.

Two traditions converging on the same structural claim is not coincidence. I view it as evidence that the claim is identifying something real about how understanding actually works: that it requires grounding in the generative structure of the world, not only in the statistical patterns of its historical outputs; that the developmental history of information is not a curiosity for philosophers and academics, but a practical determinant of what analytical systems can and cannot know; that the difference between an analysis that preserves ontogenic context and one that strips it is not a difference of degree but of kind.

The most consequential analytical failures of the past decade were not failures of data quantity, computational power, or domain expertise. They were failures of ontogenic grounding. I believe that the structural conditions that made past failures foreseeable may be encoded in the developmental history of the relevant information systems. Part of their network topology, in their coupling mechanisms, in the spatial pattern of their generative concentrations. That history was not analyzed, because the analytical frameworks in use treated the data as static resources rather than as developmental trajectories.

It follows that the most important structural relationships do not respect disciplinary boundaries. They operate through the ontogenic transfer pathways that connect information systems across those boundaries. The food-water-energy-infrastructure woven throughout the framework of this paper is a zone of dense inter-system coupling: information born in the climate system develops into agricultural meaning, then into energy meaning, then into infrastructure risk, then into financial exposure, through a chain of ontogenic transformations that no single discipline can trace in its entirety.

> *Information Ontogeny as Infrastructure is not a model. It is a way of knowing - a commitment to reading the developmental history of data as primary, and to understanding the world as a system of information systems whose couplings are the primary object of analysis. It is, I believe, the right way to know the world we are actually in.*

Selected References

Primary intellectual anchors for the arguments developed in this paper. A full bibliography accompanies the journal submission.

Source	Role in This Paper's Argument
Oyama, S. (1985 / 2000). The Ontogeny of Information: Developmental Systems and Evolution. Duke UP.	Foundational intellectual origin. Information as developmental product of organism-environment interaction, not pre-formed content residing in a medium. The direct predecessor of this paper's central claim.
Li, F.-F. (2025). From Words to Worlds: Spatial Intelligence is AI's Next Frontier. Substack / World Labs Manifesto.	Convergent validation from AI and computer vision: reasoning is constitutively dependent on perceptual grounding. The engineering instantiation of the paper's epistemological claim at the AI system level.
Li, F.-F. (2025). Spatial Intelligence Is AI's Next Frontier. TIME.	Spatial intelligence as the scaffolding of cognition; the perception-reasoning dependency; AGI without spatial intelligence is incomplete.
Shannon, C. & Weaver, W. (1949). The Mathematical Theory of Communication. University of Illinois Press.	Information theory foundations. This framework extends Shannon by insisting that context is constitutive of meaning, not separable from the signal.
Massey, D. (2005). For Space. Sage Publications.	Space as relational rather than absolute; always under construction through interactions between entities. Ontological foundation for the spatial dimension of information ontogeny.
Meadows, D. (2008). Thinking in Systems. Chelsea Green Publishing.	Systems dynamics; feedback loops; the primacy of relational structure over component properties in determining system behavior.
Batty, M. (2013). The New Science of Cities. MIT Press.	Space of flows; network science applied to cities as information systems; information as constitutive of spatial structure.
Liu, J. et al. (2013). Framing sustainability in a telecoupled world. Ecology and Society, 18(2).	Telecoupling framework: sending systems, receiving systems, spillover systems, coupling mechanisms. Structural template for supply chain teleconnection analysis.
Boers, N. et al. (2019). Complex networks reveal global pattern of extreme-rainfall teleconnections. Nature, 566.	Network methods for identifying teleconnections from structural coupling rather than statistical correlation. Template for the durable predictive power argument.
Fernandez, N. et al. (2014). Information measures of complexity, emergence, self-organization. Springer.	Information-theoretic formalization of emergence, self-organization, and complexity at multiple scales.

Holland, J. (1998). Emergence: From Chaos to Order. Addison-Wesley.	Emergence in complex adaptive systems; why component-level analysis cannot access system-level behavior. Foundation for the structural concentration of risk argument.
West, G. (2017). Scale. Penguin Press.	Universal scaling laws across biological, urban, and economic systems; structural analysis that holds across scales of organization.
Anderson, C. (2020). Complex spatial networks: Theory and geospatial applications. Geography Compass.	Review of the scholarly gap between GIS, complexity science, and network science that this paper addresses.
USP Medicine Supply Map (2024–25). Geographic concentration of API manufacturing.	Primary empirical anchor: pharmaceutical supply chain spatial concentration as self-organized structural risk legible from network topology before it manifested as operational disruption.
Chester, M. & Allenby, B (2024). Infrastructure first principles for the Anthropocene.	Describes the decoupling between traditional infrastructure development and the modern needs of today's increasingly complex environments.
Hidalgo, Ceasar (2015). Why Information Grows: The Evolution of Order, from Atoms to Economies.	*Why Information Grows* explores the origins of physical order and economic growth, expressed through an information lens.

Selected references with roles in the paper's argument.

This working paper is an evolving document. The author welcomes critical engagement, methodological challenge, collaboration and productive disagreement, which is how understanding advances at the boundaries of ideas.