# Ascending Pearl's Ladder

Operationalizing the Causal Hierarchy in Financial Modeling

Research
Aleatoric Systems

# Abstract

The integration of correlation-based artificial intelligence (AI)... has created significant systemic risk challenges, exposed model brittleness, and highlighted a lack of causal understanding. This has been linked to significant model failures (e.g., TerraUSD collapse) and has contributed to an increasing regulatory focus on "meaningful explainability"... which may render opaque "Black Box" models sub-optimal for critical regulatory functions. This report advocates for Structural Causal AI as a robust, auditable framework to meet these challenges, moving financial modeling from Level 1 (Association) to Level 3 (Counterfactuals) of Pearl's Causal Hierarchy. This architecture is built on three pillars:

1. **Structural Causal Models (SCMs):** Utilizing frameworks like **FinCARE**, which fuses LLM-extracted causal knowledge with statistical algorithms to create transparent, auditable causal graphs that provide verifiable logic for regulatory compliance.
2. **Physics-Informed Neural Networks (PINNs):** Embedding financial laws (e.g., no-arbitrage) directly into the model's loss function. Advanced variants, **RRaPINNs**, minimize the **Conditional Value-at-Risk (CVaR)** of model residuals to ensure structural integrity and robust performance in high-volatility "tail risk" scenarios. PINNs also solve the **"Startup Dilemma"** by enabling competitive modeling with minimal data.
3. **Temporal Conformal Prediction (TCP):** A distribution-free framework that provides **guaranteed uncertainty** and prediction intervals that dynamically adapt to non-stationary market regimes, proving superior to traditional risk metrics like VaR during the COVID-19 stress test.

Structural Causal AI synthesizes these elements to create **"Glass Box"** systems, shifting the industry from modeling the shadows of the market (correlation) to modeling its core machinery (causation).

# 1. The Epistemic Crisis in Modern Financial Intelligence

The integration of artificial intelligence into the global financial architecture has precipitated a profound epistemic crisis, characterized by a dangerous divergence between the perceived competence of algorithmic systems and their actual structural reliability. While the proliferation of deep learning agents and Large Language Models (LLMs) has delivered superficial efficiencies in information synthesis and automated decision-making, it has simultaneously introduced a class of systemic risks that are poorly understood and difficult to quantify using traditional metrics. This report argues that the era of unconstrained data mining is ending, driven by a convergence of catastrophic model failures during recent market crises and a stringent new global regulatory regime demanding "meaningful explainability."

The central thesis of this analysis is that the current generation of AI models—built primarily on the statistical correlation of historical data—suffers from a fundamental "epistemic gap." These models function as what recent scholarship describes as "stochastic parrots" or engines of "Potemkin interpretation"1. They generate outputs that possess the façade of expert reasoning but lack any grounding in the causal mechanisms that govern financial markets. In high-stakes environments, such as capital determination, credit underwriting, and systemic risk monitoring, this lack of understanding is not merely a technical limitation; it is a source of endogenous risk that threatens the stability of the financial system itself.

## 1.1 The Illusion of Competence: "Potemkin Interpretation" and Brittleness

Current AI models, particularly Generative AI and deep neural networks, exhibit a phenomenon best described as "superficial fluency." These systems produce outputs—whether code, market analysis, or risk assessments—that appear correct and authoritative but are often "brittle and frequently arbitrary"1. This brittleness is empirically demonstrated by the phenomenon of "prompting instability," where minor, semantically irrelevant perturbations to an input prompt can lead to diametrically opposed outputs. For instance, research indicates that changing a prompt from asking a model to "explore" a company's financial distress to asking it to "delve" into the same data can cause significantly different risk classifications[1].

This introduces a level of stochastic noise that is fundamentally incompatible with fiduciary responsibility. If a model's output is contingent on the arbitrary phrasing of a query rather than the fundamental economic reality of the asset, it cannot serve as a basis for robust risk management. This "reliability gap" is exacerbated by the "epistemic gap": the uncertainty regarding what these models are actually measuring[3]. Proponents of generative interpretation in law and finance argue that LLMs can replace human judgment in interpretive tasks. However, empirical studies show that these models do not "reason" in any legal or economic sense; they traverse a probabilistic map of language patterns. This distinction is critical because, unlike a human analyst whose errors can be audited against a logic framework, an LLM's error is often a hallucination—a plausible-sounding falsehood derived from statistical noise.

## 1.2 Systemic Feedback Loops and Endogenous Risk

The danger of correlation-based AI extends beyond individual model failure to the creation of endogenous feedback loops. When multiple algorithmic agents operate on similar flawed correlations, they can amplify market signals, detaching prices from fundamentals and precipitating liquidity crises. The collapse of the **TerraUSD (UST)** stablecoin serves as a potent case study of this "model risk"1.

Unlike traditional market crashes driven by exogenous shocks (e.g., a pandemic or geopolitical event), the Terra collapse was a mechanism design failure exacerbated by algorithmic feedback. The system relied on an arbitrage mechanism between the stablecoin (UST) and its collateral token (LUNA) to maintain a peg. However, this mechanism presupposed a level of market liquidity and uncorrelated behavior that evaporated under stress. As the peg broke, algorithmic trading bots and smart contracts, acting on pre-programmed correlation assumptions, executed a cascade of sell orders that hyper-inflated the LUNA supply, driving its value to zero[4].

This event highlights the critical distinction between **exogenous risk** (risk from the world) and **endogenous risk** (risk generated by the system itself). "Black Box" models are inherently blind to endogenous risk because they are trained on historical data where these feedback loops may not have been active or visible. They assume the market is a static environment to be predicted, rather than a dynamic system that reacts to their own predictions. This report posits that **Structural Causal AI** is the necessary evolution to address this blindness. By explicitly modeling the causal mechanisms—the "physics" of the market—risk managers can simulate counterfactual scenarios where liquidity dries up or feedback loops activate, moving beyond the limitations of historical correlation.

## 1.3 The Economic Barrier: The "Startup Dilemma"

Beyond the systemic and epistemic risks, the reliance on deep learning creates a debilitating economic barrier for new market entrants, referred to here as the **"Startup Dilemma"∗∗1. Deep learning models are notoriously data-hungry, often requiring millions of labeled examples to converge to a useful solution. The cost of acquiring, cleaning, and annotating financial data is prohibitive. Expert annotation for financial datasets—such as labeling distressed assets in balance sheets or classifying complex derivatives—requires domain expertise that commands high hourly rates, often ranging from $6 to $12 per hour or significantly higher for specialized tasks[2].

This economic structure entrenches incumbents—large banks and major technology firms—who possess proprietary data moats. Startups attempting to compete using standard deep learning architectures face a "Cold Start" problem: they cannot afford the data required to train a competitive model, and without a competitive model, they cannot acquire the customers necessary to generate data. This stifles innovation and homogenizes the modeling approaches used in the market, further increasing systemic correlation. As we will detail in Section 6, **Physics-Informed Neural Networks (PINNs)** offer a solution to this dilemma by substituting data with *structure*. By embedding financial laws directly into the learning process, PINNs can solve complex pricing and risk problems with sparse or even zero labeled data, dramatically lowering the barrier to entry and enabling a more diverse ecosystem of models[2].

# 2. The Regulatory Pivot: From Passive Guidance to Active Enforcement

The regulatory landscape for Artificial Intelligence in finance has undergone a phase transition. We have moved from an era of high-level principles and passive guidance to one of active enforcement and specific, stringent requirements. Regulatory bodies globally are converging on a single mandate: **opacity is a liability**.

## 2.1 The SEC and the Crackdown on "AI Washing"

In the United States, the Securities and Exchange Commission (SEC) has aggressively targeted **"AI washing"**—the practice of making unsubstantiated claims about an algorithm's capabilities. In March 2024, the SEC settled charges against two investment advisers, **Delphia (USA) Inc.** and **Global Predictions Inc.**, for making false and misleading statements about their use of AI5.

- **Delphia** claimed to use "predictive algorithmic models" and "machine learning" to analyze client data for investment decisions. Specifically, they marketed that they used client data from social media, banking, and credit cards to "make intelligent investment decisions." The SEC found these claims to be false; the firm was not using the touted AI capabilities in its investment process in the manner described.
- **Global Predictions** marketed itself as the "first regulated AI financial advisor" and claimed to use "Expert AI-driven forecasts." The SEC found these representations to be unsubstantiated and penalized the firm for failing to produce documentation supporting these claims.

These enforcement actions, resulting in substantial civil penalties ($225,000 for Delphia and $175,000 for Global Predictions), signal a new reality: **Model failure is a legal event.** Financial institutions must be able to substantiate their claims about AI performance. If a model is a "Black Box" whose operations cannot be verified or explained, any marketing claim regarding its "expert" nature or "predictive power" carries significant regulatory risk. The SEC has made it clear that "claims about prospects should have a reasonable basis and investors should be told that basis"6. This effectively mandates a level of transparency that pure "Black Box" models cannot provide.

### 2.2 Global Mandates for "Meaningful Explainability"

International standard-setting bodies are reinforcing this stance by tightening the definition of "explainability." It is no longer sufficient to provide a "feature importance" chart (like SHAP values) that merely indicates correlations. Regulators are demanding explanations that reveal the *causal logic* of the decision.

- **Bank for International Settlements (BIS):** The BIS has identified opaque models as a direct prudential concern. In its reports on AI in the financial sector, The BIS emphasizes that a model whose results lack auditable transparency or reproducibility presents a high prudential concern for critical business areas... The BIS has indicated that "supervisory authorities may have reduced confidence in the results of an AI model that lacks sufficient explainability."
- **International Association of Insurance Supervisors (IAIS):** The IAIS has defined **"meaningful explanations"** as providing "understandable, transparent, and relevant insights" into the decision-making process1. This definition pushes beyond technical transparency to "mental model alignment"—the explanation must make sense to a human domain expert. An explanation that relies on high-dimensional vector math is not "meaningful" to a loan officer or a regulator.
- **U.S. Regulatory Agencies:** The definition of explainability is coalescing around "how an AI approach uses inputs to produce outputs"1. This implies a mechanistic understanding, not just a statistical one.

This regulatory environment creates a binary outcome for financial firms: build **"Glass Box"** models that are inherently explainable and structurally sound, or risk being locked out of high-value enterprise markets and facing enforcement actions. The "move fast and break things" era for financial AI is effectively over; the new era demands "move deliberately and prove it."

## 3. Theoretical Framework: Ascending the Causal Hierarchy

To meet these regulatory and epistemic challenges, financial modeling must transcend "objective metrics" of accuracy (like Mean Squared Error or AUC) and adopt **Structural Fidelity** as the primary standard. This requires a theoretical framework that distinguishes between different types of information. We utilize the **Causal Hierarchy Theorem (CHT)**, formalized by Judea Pearl and recently adapted for financial contexts[1].

## 3.1 The Causal Hierarchy (Pearl's Ladder)

The CHT posits that causal information is structured in three distinct levels. Information at a lower level cannot answer questions at a higher level without additional structural assumptions. This theorem proves that no amount of Level 1 data (correlation) can produce Level 2 or Level 3 knowledge without a model of the underlying mechanism.

**Level 1: Association (Seeing)**

- **Question:** "What if I see X?" (e.g., "How does the stock price correlate with interest rates?")
- **Current State:** This is the domain of standard Machine Learning and Large Language Models. These systems are excellent at pattern recognition and curve fitting.
- **Limitation:** They are prone to the **"Correlation-Causality Fallacy."** For example, standard linear metrics often miss significant nonlinear causality between markets, leading risk managers to underestimate true asset connectivity during a crisis[1]. A model might learn that "umbrella sales" correlate with "traffic accidents" (both caused by rain) and erroneously predict that banning umbrellas will reduce accidents. In finance, this manifests as models learning spurious correlations that vanish during regime shifts.

**Level 2: Intervention (Doing)**

- **Question:** "What if I do X?" (e.g., "What will happen to liquidity if the central bank raises rates?" or "What happens if we liquidate this position?")
- **Requirement:** Answering this requires a model of the mechanism, not just a historical pattern. It requires understanding the direction of causality.
- **Financial Relevance:** This is critical for policy making and active portfolio management. It distinguishes between a passive observation and an active change in the system[1].

**Level 3: Counterfactuals (Imagining)**

- **Question:** "What if X had been different?" (e.g., "What would our portfolio value be today if we had hedged yesterday?" or "Would this borrower have defaulted if they had been given a lower interest rate?")
- **Requirement:** This requires a fully specified Structural Causal Model (SCM) that can simulate worlds that never occurred.

- **Financial Relevance:** This is the core of rigorous risk management, stress testing, and liability management. It allows firms to test their resilience against "Black Swan" events that have not yet happened but are structurally possible[1].

The proposed **Structural Causal AI** standard explicitly targets **Level 2 and Level 3**. By modeling the underlying mechanisms (the "structure"), these systems can answer interventional and counterfactual questions, providing the "meaningful explainability" regulators demand.

# 4. Pillar 1: Structural Causal Models (SCMs) for Verifiable Logic

The first pillar of the proposed architecture addresses the flaw of **logic**. Structural Causal Models (SCMs) provide a framework where the relationships between variables are defined by directed edges in a graph, representing causal influence rather than mere statistical association.

## 4.1 The FinCARE Framework: A Hybrid Discovery Pipeline

Constructing SCMs for finance is challenging due to the complexity and dimensionality of the system. We propose a hybrid pipeline, validated by the **FinCARE (Financial Causal Analysis with Reasoning and Evidence)** framework, which integrates Large Language Models (LLMs) with statistical causal discovery algorithms to overcome the limitations of each11.

The Methodology:

The FinCARE framework operates on the premise that while statistical algorithms (like PC or GES) are rigorous, they struggle with "weak signals" and often produce Markov Equivalence Classes (multiple graphs that fit the data equally well) rather than a unique causal structure. Conversely, LLMs possess vast domain knowledge but are prone to hallucination. FinCARE fuses these by using the LLM to generate a "prior" structure that constrains the statistical search.

1. **Causal Relationship Extraction:** An LLM processes vast corpora of unstructured financial documents (e.g., SEC 10-K filings, earnings call transcripts). It extracts potential causal triplets, such as (Company A) -\> -\> (Company B) or (Interest Rates) -\> \[Negatively_Impacts\] -\> (Housing Starts)1.
2. **Knowledge Graph (KG) Construction:** These extractions are assembled into a **Financial Knowledge Graph (FinReflectKG)**. Crucially, edges are scored based on a composite metric of **Strength** (confidence of extraction), **Frequency** (mention count), and **Coverage** (cross-validation across sources)[12].

3. **Constraint Injection:** These scored relationships are injected as *constraints* into statistical causal discovery algorithms.

- High-confidence links become **"Required Edges"** (the model *must* consider them).
- Illogical or biologically/economically impossible links become **"Forbidden Edges"** (the model *cannot* use them).
- This dramatically reduces the search space for the algorithm, preventing it from identifying spurious correlations (e.g., correlating stock prices with sunspots).

Empirical Validation:

The FinCARE study demonstrates that this hybrid approach drastically outperforms traditional statistical methods in recovering the true causal graph of financial networks. By using KG constraints:

| Algorithm | Base F1 Score | Constrained F1 Score | Improvement |
|---|---|---|---|
| PC Algorithm | 0.459 | 0.622 | 36% |
| GES Algorithm | 0.367 | 0.735 | 100% |
| NOTEARS | 0.163 | 0.759 | 366% |

These results validate that domain knowledge, when encoded structurally, allows algorithms to "see" causal links that are statistically weak but economically significant[12].

## 4.2 The Regulatory Payoff: Auditable Explanations

This approach solves the explainability problem by design. The model's reasoning is no longer a "black box" of weights; it is a transparent causal graph. An output looks like: *"The model predicts a revenue decline for Company A because its primary supplier, Company B, is facing a strike (verified by news reports) and there is a verified causal dependency."* This directly satisfies the IAIS and BIS requirements for "meaningful explanations" because the logic maps to real-world entities and relationships, not abstract vectors[1].

# 5. Pillar 2: Physics-Informed Neural Networks (PINNs) for Structural Integrity

The second pillar addresses the **"Startup Dilemma"** (data inefficiency) and the lack of **structural integrity**. Financial markets, like physical systems, are governed by fundamental laws (e.g., no-arbitrage conditions, option pricing PDEs). **Physics-Informed Neural Networks (PINNs)** embed these laws directly into the neural network's loss function.

## 5.1 Mechanics of PINNs in Finance

A standard neural network minimizes a data-driven loss ($\mathcal{L}data$)*, typically the Mean Squared Error between predictions and observed labels. A PINN adds a "Physics Loss" (*$\mathcal{L}{PDE}$) derived from the governing differential equation (e.g., the Black-Scholes PDE for option pricing).

The total loss function becomes:

$$\mathcal{L}total = \mathcal{L}data + \lambda \mathcal{L}_{PDE}$$

where $\mathcal{L}_{PDE}$ measures the "residual"—the degree to which the network's prediction violates the known financial law.

**Advantages for Financial Modeling:**

1. **Data Efficiency:** The model is "born" knowing the rules of finance. Research shows PINNs can solve option pricing problems with **zero** or very sparse labeled data2. This effectively bypasses the "Cold Start" problem for startups, as they do not need to purchase expensive historical option data to train a valid pricer.
2. **Structural Consistency:** The model is constrained to output prices that are consistent with no-arbitrage principles. A standard ML model might predict a negative option price or an arbitrage opportunity due to overfitting noise; a PINN is penalized heavily for such violations, ensuring "physical" plausibility[15].
3. **Computational Speed:** Once trained, a PINN acts as a "neural surrogate" that is orders of magnitude faster than traditional numerical solvers (like Finite Difference Methods or Monte Carlo simulations). For high-dimensional problems, PINNs avoid the "curse of dimensionality" that plagues grid-based methods[16].

## 5.2 Managing Tail Risk: Residual Risk-Aware PINNs (RRaPINNs)

A critical limitation of standard PINNs (and standard ML) is that they minimize *average* error (MSE). In finance, risk is concentrated in the tails—the extreme events. A model that is accurate on average but fails during a market crash is useless for risk management. Standard PINNs can exhibit "propagation failure," where the solution is accurate in the bulk of the domain but violates the PDE in critical, high-gradient regions (e.g., near the strike price at maturity)[17].

To address this, we advocate for **Residual Risk-Aware PINNs (RRaPINNs)**. This advanced architecture changes the optimization objective from minimizing MSE to minimizing the **Conditional Value-at-Risk (CVaR)** of the residuals[17].

Methodology:

The RRaPINN formulation replaces the standard loss with a risk-averse objective:

$$\min_{\theta} \mathrm{CVaR} \alpha(|\mathcal{R}(u\theta)|)$$

where $\mathcal{R}$ is the PDE residual and $\alpha$ is the confidence level (e.g., 95%). Because CVaR is coherent and convex, it provides dense, informative gradients even from the tail of the distribution.

To improve optimization stability, RRaPINNs often utilize a Mean-Excess (ME) surrogate penalty. This penalty explicitly targets the worst-case errors—the "tail residuals" that occur in high-volatility regions—by penalizing the positive excess of the residual tail beyond an adaptive tolerance $\epsilon$[17].

**Impact:**

This forces the model to focus its learning capacity on the "hardest" parts of the problem (e.g., deep out-of-the-money options or regime shifts). Empirical benchmarks on PDEs show that RRaPINNs significantly reduce the maximum error ($L_\infty$ norm) and the tail distribution of errors compared to standard PINNs. This ensures the model remains robust even in "Black Swan" scenarios, making it suitable for stress-testing liquidation engines[17].

## 5.3 Application: American Option Pricing

Pricing **American options** (which can be exercised early) is computationally expensive because it involves a "free-boundary" problem. The boundary between the region where you should hold the option and the region where you should exercise it is unknown and moves over time. Traditional methods (like Finite Difference) require dense grids and are computationally intensive.

PINNs solve this by treating the early exercise condition as an inequality constraint in the loss function. Recent studies confirm that PINNs can accurately price American options and compute Greeks (sensitivities like Delta and Gamma) efficiently via automatic differentiation. This offers a mesh-free global solution valid for any asset price and time-to-maturity[14].

By utilizing PINNs, financial institutions can obtain pricing models that are not only faster but also mathematically guaranteed to respect the boundary conditions of the contract, providing a level of structural assurance that purely data-driven models cannot match.

# 6. Pillar 3: Conformal Prediction (CP) for Guaranteed Uncertainty

The third pillar addresses the **unreliable uncertainty estimation** of "Black Box" models. Risk management requires not just a point prediction (e.g., "The stock will be $100"), but a rigorous interval (e.g., "The stock will be between $95 and $105 with 95% probability"). Traditional methods like Value-at-Risk (VaR) rely on distributional assumptions (e.g., normality) that are frequently violated in real-world financial markets, leading to catastrophic underestimation of risk during crises.

## 6.1 The Validity Problem: Exchangeability vs. Non-Stationarity

**Conformal Prediction (CP)** is a distribution-free framework that provides **finite-sample coverage guarantees**. A 95% conformal prediction interval is mathematically guaranteed to contain the true value 95% of the time, regardless of the underlying distribution of the data (Normal, heavy-tailed, skewed, etc.)[19].

However, standard CP relies on the assumption of **exchangeability**—that the data points are drawn from the same distribution and their order does not matter. Financial time series violate this assumption; they are **non-stationary**, exhibiting volatility clustering, regime shifts, and trends. Applying standard CP to financial data often results in intervals that lose validity during volatility spikes (under-coverage) or become inefficiently wide during calm periods (over-coverage)[20].

## 6.2 The Solution: Temporal Conformal Prediction (TCP)

To resolve this, we introduce **Temporal Conformal Prediction (TCP)**, specifically variants like **TCP-RM** (Robbins-Monro) or Adaptive Conformal Inference (ACI). These methods are designed for non-exchangeable, sequential data.

Mechanism:

TCP employs an online feedback loop to dynamically calibrate the width of the prediction interval. It monitors the "coverage error" (did the previous interval cover the true value?) at each time step.

- If the model *under-covers* (errors are too frequent), the algorithm increases the scaling factor, widening the next interval.
- If the model *over-covers* (intervals are too conservative), it decreases the scaling factor, shrinking the interval.

This "adaptive calibration" allows the model to react to changing market regimes in real-time20.

## 6.3 Empirical Evidence: The COVID-19 Stress Test

The superiority of TCP was starkly demonstrated during the **COVID-19 market crash** (March 2020), a period of extreme volatility that broke many traditional risk models. A benchmarking study compared TCP against **GARCH** (a standard econometric volatility model) and **Historical Simulation** (a standard VaR approach) across three asset classes: S&P 500, Bitcoin, and Gold20.

Results Summary (Target Coverage: 95%)

| Asset | Model | Empirical Coverage | Average Interval Width | Behavior During Crisis |
|---|---|---|---|---|
| **S&P 500** | **TCP** | **95.2%** | 5.21 | Intervals widened *immediately* at volatility onset. |
| | GARCH | 82.7% | 3.05 | Failed to adapt; consistently under-estimated risk. |
| | Hist. Sim. | 93.1% | 5.06 | Reacted with a lag; slow to capture the spike. |
| **Bitcoin** | **TCP** | **95.4%** | 20.89 | Maintained valid coverage despite extreme crypto volatility. |
| | GARCH | 85.3% | 11.39 | Dangerous under-coverage. |

Implications:

The GARCH model produced "sharper" (narrower) intervals, which might appear more precise to a naive user. However, this sharpness came at the cost of validity; it demonstrated an empirical under-coverage relative to the target rate in this study's test period. In a leverage-constrained environment, this underestimation leads to margin calls and liquidation cascades.

TCP, by contrast, prioritized validity. It sacrificed sharpness (wider intervals) to maintain the 95% coverage guarantee. During the crash, TCP intervals "inflated" instantly to capture the volatility, mitigating the risk of extreme underestimation during stress events. This research indicates that TCP provides a statistically more robust standard for regulatory capital calculations, where the primary goal is solvency and safety, not just precision[20].

## 7. Case Study: Mitigating Risk in Fintech Lending

To demonstrate the practical necessity of this framework, we examine the failure modes of legacy Fintech lending models during the COVID-19 pandemic and how Structural Causal AI provides a remedy. This sector serves as a microcosm for the broader "data vs. structure" debate.

### 7.1 Failures of Correlation-Based Pricing

Fintech lenders entered the market with the promise of using "Big Data" and advanced algorithms to price risk more accurately than traditional banks. However, NBER research reveals that during the stress of the pandemic, these models reverted to simple, brittle correlations.

- **Pricing Failure:** Instead of leveraging their advanced data to distinguish risk, Fintech lenders charged a massive **45% premium** on interest rates for nonprime borrowers compared to prime borrowers with similar default risk21. This inefficiency was driven by an over-reliance on traditional **FICO scores** as a "catch-all" proxy for risk. The algorithms, unable to process the structural break of the pandemic, defaulted to a coarse heuristic: "low FICO = high risk," ignoring other predictive variables that remained stable.
- **Constraint Failure:** Algorithms failed to account for structural constraints like **funding liquidity**. Fintech lenders often rely on securitization markets to fund loans. When these markets tightened, lenders indiscriminately cut credit supply. The models were trained on borrower-level correlations and were blind to this systemic, lender-level constraint[1].

## 7.2 The Structural Solution: RappiCard and Alternative Data

A contrasting study of **RappiCard** in Mexico demonstrates the power of Structural Causal AI using **alternative data** to bridge the credit gap for unbanked populations[22].

- **The Problem:** Most applicants in Mexico lack a credit bureau history ("thin-file"), making them invisible to traditional FICO-based models.
- **The Structural Approach:** RappiCard utilized a causal model based on digital transaction history from its delivery app (e.g., order frequency, tips, payment velocity). Instead of relying on a proxy like FICO, the model identified a causal pathway:
  Transaction Volume -\> Cash Flow Velocity -\> Repayment Capacity
- **Results:** The machine learning model using this alternative data achieved an **AUC (Area Under the Curve) of 0.752**, significantly outperforming traditional metrics.
- **Fairness Intervention:** By applying a **gender-segmented** structural model (recognizing that women and men may have different causal drivers for repayment due to socioeconomic factors), the lender improved fairness. The segmented model approved **12.3%** of women who would have been rejected by a pooled (gender-blind) model, *without* increasing the default rate[22].

**Implication:** A structural approach allows lenders to disentangle *ability to repay* from *historical bias* (lack of credit history). By mapping the causal structure of cash flows, lenders can safely extend credit to underserved populations, solving both a business problem (market expansion) and a regulatory problem (fair lending).

# 8. Conclusion: The "Glass Box" Standard

The reliance on data-hungry, correlation-based "Black Box" AI is no longer a viable strategy for modern finance. It is epistemically fragile, legally dangerous, and economically inefficient. The "illusion of competence" provided by LLMs and standard deep learning models shatters under the stress of non-stationary markets and rigorous regulatory scrutiny.

This report validates **Structural Causal AI** as the new gold standard for model risk management. By synthesizing the three pillars detailed in this analysis, financial institutions can engineer **"Glass Box"** systems that meet the demands of the modern era.

1. **Structural Causal Models (SCMs)** provide **verifiable logic**. Through frameworks like FinCARE, firms can map the causal pathways of risk (Level 3 Causality), ensuring that model outputs are driven by economic reality rather than spurious correlation. This directly satisfies the "meaningful explainability" mandates of the BIS and SEC.

2. **Physics-Informed Neural Networks (PINNs)** provide **structural integrity**. By embedding financial laws into the loss function, and specifically utilizing **RRaPINNs** to minimize tail risk (CVaR), firms can build models that are robust to "Black Swan" events and data-efficient enough to solve the "Startup Dilemma."
3. **Temporal Conformal Prediction (TCP)** provides **guaranteed uncertainty**. By replacing heuristic measures like VaR with mathematically valid prediction intervals that adapt to non-stationarity, firms can ensure solvency during crises, as evidenced by TCP's superior performance during the COVID-19 crash.

For incumbents and startups alike, the transition to Structural Causal AI is not an option—it is an existential imperative. It represents the shift from modeling the *shadows* of the market (correlation) to modeling the *machinery* of the market (causation)

c

---

## 9. List of Abbreviations

- **AI:** Artificial Intelligence
- **AUC:** Area Under the Curve (Receiver Operating Characteristic)
- **BIS:** Bank for International Settlements
- **CHT:** Causal Hierarchy Theorem
- **CP:** Conformal Prediction
- **CVaR:** Conditional Value-at-Risk
- **GARCH:** Generalized Autoregressive Conditional Heteroskedasticity
- **IAIS:** International Association of Insurance Supervisors
- **KG:** Knowledge Graph
- **LLM:** Large Language Model
- **ME:** Mean-Excess (Penalty)
- **MSE:** Mean Squared Error
- **PDE:** Partial Differential Equation
- **PINN:** Physics-Informed Neural Network
- **RRaPINN:** Residual Risk-Aware PINN
- **SCM:** Structural Causal Model
- **SEC:** U.S. Securities and Exchange Commission
- **TCP:** Temporal Conformal Prediction

- **VaR:** Value-at-Risk

Works cited

1. AleatoricWhitePaper
2. AI Modeling for Risk and Prediction.docx
3. Generative Misinterpretation - James Grimmelmann, accessed December 18, 2025, https://james.grimmelmann.net/files/articles/generative-misinterpretation.pdf
4. An Econometric and Time Series Analysis of the USTC Depeg's Impact on the LUNA Classic Price Crash During Spring 2022's Crypto Market Turmoil - MDPI, accessed December 18, 2025, https://www.mdpi.com/2813-2432/3/4/24
5. Securities and Exchange Commission Brings First Enforcement Actions Over "AI-Washing" | Insights | Mayer Brown, accessed December 18, 2025, https://www.mayerbrown.com/en/insights/publications/2024/04/securities-and-exchange-commission-brings-first-enforcement-actions-over-aiwashing
6. SEC Targets "AI Washing" by Companies, Investment Advisers, and Broker-Dealers, accessed December 18, 2025, https://www.winston.com/en/blogs-and-podcasts/capital-markets-and-securities-law-watch/sec-targets-ai-washing-by-companies-investment-advisers-and-broker-dealers
7. BIS Report says regulatory action on AI model explainability is imperative, accessed December 18, 2025, https://www.grip.globalrelay.com/bis-report-says-regulatory-action-on-ai-model-explainability-is-imperative/
8. Managing explanations: how regulators can address AI explainability - Bank for International Settlements, accessed December 18, 2025, https://www.bis.org/fsi/fsipapers24.pdf
9. A Survey of Methods, Challenges and Perspectives in Causality - arXiv, accessed December 18, 2025, https://arxiv.org/html/2302.00293v3
10. On Pearl's Hierarchy and the Foundations of Causal Inference - ResearchGate, accessed December 18, 2025, https://www.researchgate.net/publication/359020799_On_Pearl's_Hierarchy_and_the_Foundations_of_Causal_Inference
11. FinCARE: Financial Causal Analysis with Reasoning and Evidence - ResearchGate, accessed December 18, 2025, https://www.researchgate.net/publication/396847491_FinCARE_Financial_Causal_Analysis_with_Reasoning_and_Evidence
12. FinCARE: Financial Causal Analysis with Reasoning and Evidence - arXiv, accessed December 18, 2025, https://arxiv.org/html/2510.20221v1

13. FinCARE: Financial Causal Analysis with Reasoning and Evidence - ChatPaper, accessed December 18, 2025, https://chatpaper.com/paper/202744
14. An Uncertainty-Aware Physics-Informed Neural Network Solution for the Black–Scholes Equation: A Novel Framework for Option Pricing - arXiv, accessed December 18, 2025, https://arxiv.org/html/2511.05519v1
15. Deep neural networks enable accurate pricing of American options under stochastic volatility | EurekAlert!, accessed December 18, 2025, https://www.eurekalert.org/news-releases/1110444
16. Finance-Informed Neural Networks — Deep Learning for Functional Problems in Macroeconomics and Finance - Carnegie Mellon University, accessed December 18, 2025, https://www.cmu.edu/sites/default/files/cmu-tepper-site-files/documents/Mott%20Dissertation.pdf
17. (PDF) RRaPINNs: Residual Risk-Aware Physics Informed Neural Networks - ResearchGate, accessed December 18, 2025, https://www.researchgate.net/publication/397933587_RRaPINNs_Residual_Risk-Aware_Physics_Informed_Neural_Networks
18. RRaPINNs: Residual Risk-Aware Physics Informed Neural Networks - arXiv, accessed December 18, 2025, https://arxiv.org/html/2511.18515v1
19. An introduction to conformal inference for economists - EconStor, accessed December 18, 2025, https://www.econstor.eu/bitstream/10419/308058/1/1913045811.pdf
20. Temporal Conformal Prediction (TCP): A Distribution-Free Statistical and Machine Learning Framework for Adaptive Risk Forecasting - arXiv, accessed December 18, 2025, https://arxiv.org/html/2507.05470
21. NBER WORKING PAPER SERIES FINTECH LENDING WITH LOWTECH PRICING Mark J. Johnson Itzhak Ben-David Jason Lee Vincent Yao Working Pa, accessed December 18, 2025, https://www.nber.org/system/files/working_papers/w31154/w31154.pdf
22. NBER WORKING PAPER SERIES FINTECH LENDING TO BORROWERS WITH NO CREDIT HISTORY Laura Chioda Paul Gertler Sean Higgins Paolina C., accessed December 18, 2025, https://www.nber.org/system/files/working_papers/w33208/w33208.pdf
23. NBER WORKING PAPER SERIES FINTECH LENDING TO BORROWERS WITH NO CREDIT HISTORY Laura Chioda Paul Gertler Sean Higgins Paolina C., accessed December 18, 2025, https://www.nber.org/system/files/working_papers/w33208/revisions/w33208.rev0.pdf

## IMPORTANT LEGAL DISCLOSURES AND DISCLAIMERS

**1. General Disclaimers and No Investment Advice.**
This document, "Ascending Pearl's Ladder—Operationalizing the Causal Hierarchy in Financial Modeling," is published by Aleatoric Systems for informational and discussion purposes only. It is intended solely for academic, research, and general market education and does not constitute a recommendation, solicitation, offer, or advice to purchase or sell any security, financial instrument, or digital asset, or to engage in any specific investment strategy. The financial, regulatory, and technical concepts discussed are complex and should not be relied upon without independent professional advice. Aleatoric Systems is not a registered investment advisor, broker-dealer, or financial institution. The views expressed are those of the author(s) and do not necessarily reflect the opinion of any other entity.

**2. Limitation of Warranty and Liability.**
The information and models described herein (including, without limitation, Structural Causal AI, SCMs, FinCARE, PINNs, RRaPINNs, and TCP) are provided "AS IS" without any warranties of any kind, express or implied. Past performance, simulated or hypothetical results, or model backtests are not guarantees or reliable indicators of future performance. Financial modeling is inherently subject to significant risks and uncertainties, and no representation is made that any account will or is likely to achieve profits or losses similar to those discussed. Aleatoric Systems expressly disclaims all liability for any direct, indirect, special, incidental, consequential, or punitive damages arising from the use of, or reliance on, this white paper.

**3. Intellectual Property and Proprietary Information.**


**4. Digital Asset Risk Notice.**
Any discussion of digital assets (e.g., stablecoins, cryptocurrencies) is for illustrative, educational, or case-study purposes only. Digital assets are highly volatile, involve a high degree of risk, and may be deemed securities in various jurisdictions, including under U.S. SEC jurisdiction, depending on the facts and circumstances. The regulatory landscape is uncertain and subject to change.

**Section: 5. Governing Law and Dispute Resolution.**
All matters concerning the interpretation, validity, and enforcement of the Intellectual Property and Proprietary Information rights claimed in Section 3 shall be governed by and construed in accordance with the laws of the State of Illinois, USA, without regard to conflict of law principles.