QUANTUMIST.AI RESEARCH SERIES | PAPER NO. 01
Operational Science of AI in Finance

THE COGNITIVE PRIMITIVES OF INVESTMENT BANKING:

An Ontology for AI-Driven Augmentation in High-Stakes Finance

Umran Nayani
Mumbai, India
December 24, 2025

Abstract

The discourse on artificial intelligence in investment banking suffers from a fundamental category error. Current analyses focus on role displacement—asking which positions AI will eliminate—rather than identifying the atomic cognitive functions that constitute financial intermediation. This paper proposes a formal ontology that decomposes investment banking into four classes of cognitive primitives, each defined by distinct truth conditions, error tolerances, and computational reducibility. We demonstrate that AI integration succeeds or fails based on primitive classification accuracy, not technological capability. By mapping the complete stack of banking cognition—from deterministic calculation to irreducibly human judgment—we establish a framework for building AI-native financial institutions that amplify rather than replace human intelligence. This ontology serves as the theoretical foundation for a forthcoming benchmark suite that will operationalize these distinctions and measure LLM performance against the true functional requirements of high-stakes finance. This paper defines the ontology and truth conditions; Papers 2–3 operationalize them into a benchmark harness and failure taxonomy for specific model evaluation.

JEL Classification:
G24 (Investment Banking; Venture Capital; Brokerage)
O33 (Technological Change: Choices and Consequences)
D83 (Search; Learning; Information and Knowledge)
L23 (Organization of Production)
C45 (Neural Networks and Related Topics)

Keywords:
Investment Banking, Artificial Intelligence, Cognitive Ontology, Large Language Models,

Task Decomposition, Operational Risk, Organizational Design, FinTech, Hallucination, Human-Computer Interaction.

I. THE CATEGORY ERROR: Why the Current Debate Is Asking the Wrong Question

In March 2024, the U.S. Securities and Exchange Commission fined two investment advisers $400,000 for "AI-washing"—falsely claiming they used artificial intelligence in their investment processes when they did not. One firm marketed itself as the "first regulated AI financial advisor". The other promised its "proprietary AI-based technology" would generate extraordinary returns with "100% protection" for client funds. Neither claim was true. Both firms settled.

This was not an isolated incident. In fiscal year 2024 alone, the SEC charged three firms for false statements about AI use, part of a broader enforcement agenda that extracted a record $8.2 billion in penalties—the highest amount in agency history. The regulator's message was unambiguous: exaggerate your AI capabilities, face consequences.

Yet these enforcement actions miss the deeper crisis. The problem is not that firms lie about using AI. The problem is that *firms don't know what using AI actually means*.

The financial services industry invested $650 billion globally in AI and data science in 2023, with spending projected to reach $1.83 trillion by 2033. Banking alone deployed $31.3 billion in AI spending in 2024, up 52% from the prior year. Over 80% of Tier-1 investment banks now run AI systems across front, middle, and back offices. This is not experimentation—it is wholesale institutional transformation.

But the capital is being incinerated: across industry surveys and practitioner analyses, a large fraction of AI initiatives fail to translate into measurable organizational value beyond pilots and experimentation. Only 33% of AI projects advance beyond experimentation; just 15-20% reach deployment; a mere 8% generate measurable business impact. Analysis by S&P Global projects a 48% failure rate on AI investments at the proof-of-concept stage. Research published in 2024 found that banks experiencing AI incidents suffer higher bankruptcy risk and lower operational cash flows compared to peers without such failures.

The pattern is clear and catastrophic: **the industry is spending hundreds of billions of dollars on a technology it has not correctly diagnosed**.

Note on figures. The quantitative figures cited above are compiled from publicly available regulator releases, industry surveys, and practitioner analyses listed in the References. Where estimates vary across sources, we intentionally report representative ranges rather than claim a single definitive point estimate.

1.1 The Flawed Question

The central debate animating boardrooms, academic conferences, and regulatory hearings remains: *"Which roles will AI replace?"* This framing is fatally misconceived.

It treats "Associate," "Vice President," and "Managing Director" as functional categories when they are, in fact, sociological artifacts—prestige markers evolved from the apprenticeship systems of 19th-century merchant banks. A second-year analyst calculating weighted average cost of capital and the same analyst consoling a CEO through a hostile takeover defense occupy the same org-chart box but perform cognitively unrelated functions. One task is deterministic computation; the other is emotional labor under radical uncertainty.

Survey data reveals why this confusion persists: 73% of financial institutions lack sufficient understanding of technical skill requirements for AI roles. Forty percent cite security and data privacy concerns as their primary AI obstacle; 33% highlight workforce skill gaps; 30% cannot measure return on investment. The industry is attempting to deploy a general-purpose technology without understanding what makes tasks suitable for that technology in the first place.

1.2 The Institutional Illusion

Investment banks are not collections of people. They are collections of *functions*—specifically, a vertically integrated stack of cognitive processes designed to:

1. **Price risk** under informational asymmetry
2. **Allocate capital** across time and uncertainty
3. **Manufacture trust** in high-stakes transactions
4. **Absorb reputational liability** as principal agents

These functions existed before the modern investment bank and will persist after it. What changes is *how* they are executed and *by whom*. The error lies in conflating the execution layer (junior bankers formatting pitch decks) with the function layer (transforming uncertainty into tradeable securities).

1.3 Why Current AI Strategies Are Failing

Despite aggressive adoption, results remain paradoxical. Algorithmic trading—one of finance's earliest AI successes—has grown into a $2.36 billion global market. AI-driven credit models demonstrate 30-40% improvement in default prediction accuracy over traditional approaches. Morgan Stanley's GPT-4-powered research assistant now synthesizes insights from 100,000+ internal documents, a task that once required reading dozens of reports manually.

Yet in adjacent domains, catastrophic failures proliferate. Large language models hallucinate with confidence, generating plausible but incorrect valuations that pass initial review.

Regulatory experts warn that AI hallucinations in financial disclosures "could result in noncompliance and legal penalties". When critical business decisions rest on flawed AI outputs without human oversight, the financial repercussions can be severe. Reputational damage follows: a bank relying on an LLM that produces false statements loses customer trust—particularly fragile in finance.

The pattern is clear: **success and failure do not correlate with task "difficulty" but with task type**. The industry lacks a taxonomy for distinguishing which cognitive primitives can be safely delegated to probabilistic systems versus which require deterministic guarantees or irreducibly human judgment. This taxonomic void produces two failure modes:

Type I Error (False Automation): Delegating probabilistic tasks to systems that hallucinate with confidence, generating plausible but incorrect outputs that explode under market stress or regulatory scrutiny.

Type II Error (False Irreducibility): Preserving human labor for tasks already computationally solved, wasting expensive expertise on mechanical work while competitors gain 30-40% efficiency advantages.

1.4 The Stakes

This is not an academic exercise. Financial institutions that correctly map cognitive primitives to AI capabilities will achieve structural cost advantages that compound into insurmountable moats. Those that misclassify will face:

- **Reputational collapse** from AI-generated errors in client-facing materials
- **Regulatory penalties** from compliance violations caused by hallucinated disclosures—as evidenced by the SEC's record enforcement agenda
- **Talent exodus** as high-performing analysts leave for firms where they are not doing work software can execute for $0.02/hour
- **Bankruptcy risk** from operational cash flow deterioration tied to AI incidents

The question is not "Will AI replace bankers?" but **"What is a bank?"** Once we answer that question at the atomic level—decomposing institutions into their constituent cognitive primitives—the path to AI integration becomes self-evident.

This paper provides that decomposition.

II. FOUNDATIONS: Task Decomposition and Computational Reducibility

Before constructing our ontology, we must establish theoretical foundations drawn from cognitive science, computer science, and organizational economics.

2.1 The Computational Theory of Mind

Human cognition can be understood as information processing over symbolic representations. This does not imply that minds *are* computers, but that both instantiate functions mapping inputs to outputs under resource constraints. For our purposes, the relevant constraint is *computational cost*—the time, memory, and energy required to execute a cognitive operation.

Research demonstrates that humans decompose complex tasks hierarchically to minimize computational cost while maximizing utility. When planning a route through a city, we do not evaluate every possible path; we identify landmark subgoals (the café, the post office) that chunk the problem space into manageable units. This hierarchical decomposition is not a bug but a feature—the rational allocation of limited cognitive resources.

Investment banking exhibits the same structure. A sell-side M&A mandate decomposes into:

- **Strategic positioning** (Should we sell now? To whom?)
- **Valuation modeling** (What is the enterprise worth?)
- **Buyer identification** (Who has strategic synergies?)
- **Marketing execution** (How do we create competitive tension?)
- **Negotiation management** (How do we maximize price?)

Each subtask decomposes further. Valuation modeling subdivides into precedent transaction analysis, discounted cash flow modeling, and leveraged buyout analysis. Each primitive operates under different truth conditions and error tolerances.

2.2 Classes of Computational Reducibility

We classify investment banking functions according to their underlying **truth conditions** and the degree to which they admit computational reduction. This classification is independent of tools, job titles, or organizational roles.

Class I — Deterministic Functions
A deterministic function is one for which a single correct output exists for a given input under shared assumptions. Correctness is evaluated by equality within specified numerical precision.

Formally, given input *x*, there exists exactly one correct output *y* such that *f(x) = y* for all observers operating under the same assumptions. Errors are discrete and binary: the output is either correct or incorrect.

Examples include arithmetic operations, financial statement aggregation, unit conversions, and the mechanical computation of ratios once inputs are specified.

Truth condition: equality
Error tolerance: zero

Class II — Probabilistic Functions
A probabilistic function produces outputs whose correctness is evaluated relative to an empirical distribution rather than exact equality. Multiple outputs may be acceptable, provided their likelihoods are properly calibrated.

Given input *x*, the output *y* is assessed by whether the implied probability distribution aligns with observed frequencies or reasonable priors. Errors are continuous rather than discrete.

Examples include forecasting revenue growth, estimating equity risk premia, or projecting default probabilities.

Truth condition: calibration
Error tolerance: continuous

Class III — Contextual Functions
A contextual function is one in which the optimal output depends on latent variables not contained in the observable input. Multiple qualitatively different outputs may be reasonable depending on constraints, incentives, timing, and strategic objectives.

High variance in outputs is acceptable; however, **systematic bias is not**. Bias here refers to persistent strategic misframing—such as consistently recommending delay, leverage, or acceleration regardless of client-specific constraints.

Examples include acquisition timing advice, capital structure recommendations, and deal strategy formulation.

Truth condition: counterfactual effectiveness
Error tolerance: high variance, low bias

Class IV — Principal (Irreducibly Human) Functions
Principal functions are those in which correctness depends not on output quality alone, but on accountable agency. Legal liability, fiduciary duty, and reputational responsibility are non-transferable properties of human or institutional principals.

These functions cannot be delegated to non-persons because legitimacy—not correctness—is the binding constraint.

Examples include signing fairness opinions, representing clients in negotiations, and assuming fiduciary responsibility.

Truth condition: legitimacy and accountability
Error tolerance: existential

2.3 Why This Taxonomy Matters for AI

Summary Table (Truth Conditions).
Class I: Truth = equality (right/wrong).
Class II: Truth = calibration (confidence matches frequencies).
Class III: Truth = counterfactual usefulness (works given latent variables + constraints).
Class IV: Truth = legitimacy (accountable agency + fiduciary authority).

Key implication: AI safety is not a property of the model in isolation, but of the model–task pairing under the relevant truth condition.

Current AI systems—particularly large language models—are universal function approximators. They can mimic any pattern in their training data, including patterns that do not generalize. This creates a dangerous symmetry: LLMs perform deterministic and probabilistic tasks with equal *apparent* competence while possessing radically different *actual* reliability.

An LLM asked to "calculate the NPV of this project" will generate text that resembles a calculation. But unless the model is equipped with a symbolic solver, it is *predicting* what the answer should look like based on pattern matching, not *computing* the answer through formal operations. The output may be correct by coincidence or catastrophically wrong with no mechanism for detecting the difference.

This is the central risk: **AI hallucination is not a surface bug; it is a structural property of probabilistic generation under uncertainty—particularly when outputs are rewarded for plausibility rather than verified correctness.** Models trained on text generate text that maximizes likelihood under their learned distribution. When the task requires deterministic correctness, probabilistic generation is the wrong tool—regardless of model size, training data, or architectural sophistication.

Conversely, for genuinely probabilistic tasks—"Generate five plausible downside scenarios for this merger"—LLMs excel precisely *because* they are sampling from learned distributions rather than executing formal logic. The error is not in the tool but in the tool-task mismatch.

Operational rule: If correctness is evaluated by equality (Class I), then any generative model output must be treated as a *proposal* until verified by a symbolic solver or deterministic computation layer. Surface plausibility is not evidence of correctness.

2.4 The Organizational Implications

If different cognitive primitives require different computational substrates, then organizational design must reflect this functional stratification. The traditional pyramid—junior bankers executing rote tasks, senior bankers making strategic decisions—emerged when *all* tasks required human cognition.

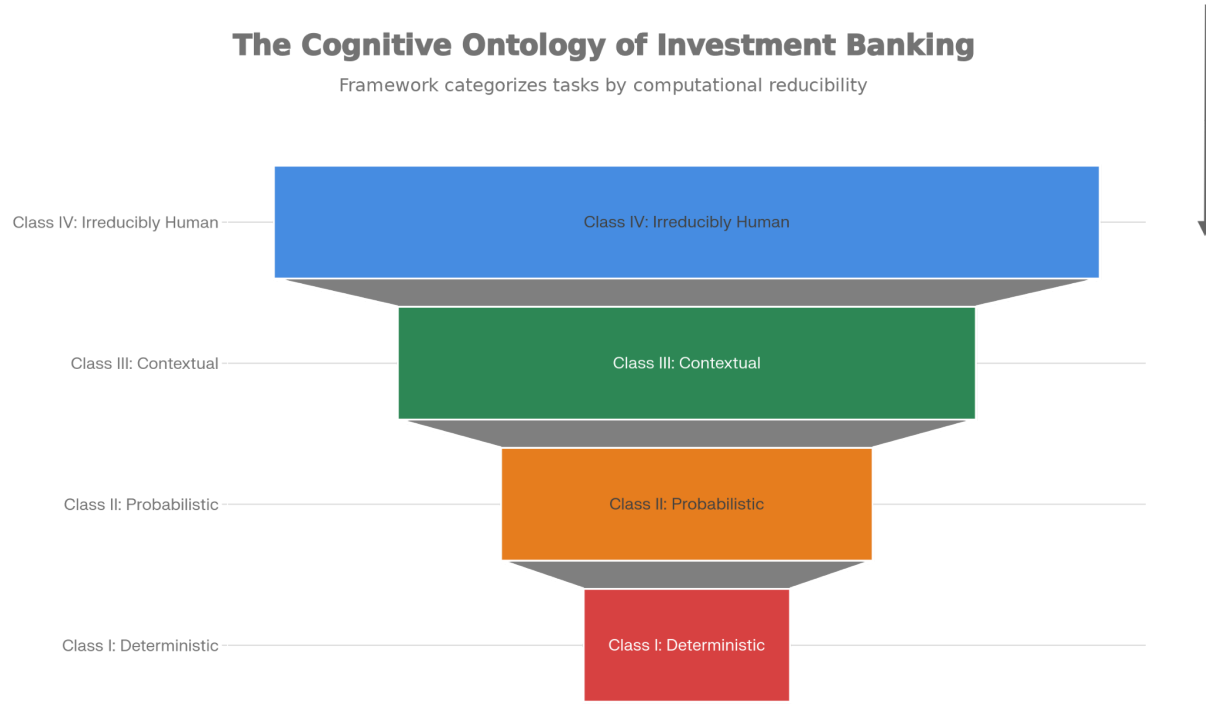
In the AI-native bank, organizational structure must align with computational architecture:

- **Deterministic tasks** migrate to symbolic solvers (zero human involvement)
- **Probabilistic tasks** utilize LLMs with human validation (human-in-the-loop)
- **Contextual tasks** employ AI as sparring partner (human-in-command)
- **Principal tasks** remain exclusively human (AI as tool, never agent)

This is not automation; it is *reorganization*. The functions persist; their execution layer transforms. Junior bankers are not "replaced"—their role evolves from executing deterministic work to validating probabilistic outputs and synthesizing contextual insights. This shift requires new skills but does not eliminate the need for human intelligence. If anything, it increases the leverage on genuine expertise while eliminating the apprenticeship model's inefficiencies.

III. THE ONTOLOGY: Mapping Investment Banking to Cognitive Primitives

We now present the core contribution: a formal ontology that maps every significant action in the investment banking value chain to one of four primitive classes. This is not a complete enumeration—that would require a multi-volume taxonomy—but a representative sample demonstrating the classification logic.



3.1 Class I: Deterministic Primitives (The "Calculator")

Definition: A cognitive operation with a single correct output for a given input under shared assumptions; correctness is evaluated by equality (within specified precision). Tasks where f(x)=y are provably correct for all observers under identical inputs. These are computationally reducible to symbolic manipulation.

Truth Condition: Correctness is binary. An NPV calculation is either accurate to the specified precision, or it is not.

Error Tolerance: Zero. Deterministic errors compound into systemic failures.

AI Role: Solver—execute deterministic operations with formal verification.

Representative Tasks:

1. **Spreading Financial Statements**
    - Input: Raw financial filings (10-K, 10-Q)
    - Output: Standardized historical financials (Income Statement, Balance Sheet, Cash Flow)
    - Computational Operation: Data extraction + arithmetic verification

- Current AI Capability: 95%+ accuracy with OCR + structured parsers
- Failure Mode: Template mismatches (e.g., non-GAAP adjustments)
- Human Role: Schema definition, exception handling

2. **Calculating WACC (Given Inputs)**
    - Input: Cost of equity, cost of debt, tax rate, capital structure
    - Output: Weighted average cost of capital
    - Computational Operation: *Weighted Average Cost of Capital (WACC) is computed mechanically as:*

      WACC = (E/V)·re + (D/V)·rd·(1 − T),

      where E and D represent equity and debt values respectively, V = E + D, re is the cost of equity, rd is the cost of debt, and T is the marginal tax rate.

    - Current AI Capability: 100% with symbolic solver
    - Failure Mode: Garbage in, garbage out (incorrect inputs)
    - Human Role: Input validation, assumption justification
3. **Formatting Pitch Decks**
    - Input: Content requirements + brand guidelines
    - Output: PowerPoint conforming to specifications
    - Computational Operation: Template application + style enforcement
    - Current AI Capability: 90%+ for standard formats
    - Failure Mode: Creative layouts requiring aesthetic judgment
    - Human Role: Initial design, final review
4. **Transaction Comparables Screening**
    - Input: Target characteristics (industry, size, timing, deal type)
    - Output: List of precedent transactions matching criteria

- Computational Operation: Database query with Boolean filters
- Current AI Capability: 100% for structured databases
- Failure Mode: Qualitative similarity (e.g., "strategic" vs "financial" buyers)
- Human Role: Defining relevance criteria, interpreting edge cases

5. **Accretion/Dilution Analysis**
    - Input: Acquirer financials, target financials, deal terms
    - Output: Pro forma EPS impact
    - Computational Operation: Arithmetic with defined formulas
    - Current AI Capability: 100% with symbolic solver + error checking
    - Failure Mode: Complex deal structures (earnouts, contingent consideration)
    - Human Role: Modeling assumptions, sensitivity analysis

Key Insight: These tasks dominated junior banker workflows until 2010. Their deterministic nature makes them the first wave of AI displacement—not because AI is "smart" but because these tasks never required intelligence. They required precision and endurance, attributes where machines have always exceeded humans.

3.2 Class II: Probabilistic Primitives (The "Predictor")

Definition: A cognitive operation whose correctness is evaluated by calibration against an empirical distribution; outputs are assessed by likelihood, not equality. Tasks where the goal is estimating **P(y|x)** or generating plausible samples from learned distributions.

Truth Condition: Calibration—does the model's confidence match empirical frequencies? Are generated samples in-distribution?

Error Tolerance: Task-dependent. Financial forecasting tolerates higher variance than fraud detection.

AI Role: Generator—produce candidates requiring human selection and validation.

Representative Tasks:

1. **Downside Scenario Generation**
    - Input: Base case financial model

- Output: Plausible stress scenarios (recession, margin compression, competitive disruption)
- Computational Operation: Pattern matching on historical crises + creative recombination
- Current AI Capability: High—LLMs excel at generating diverse, plausible scenarios
- Failure Mode: Overconfident predictions; missing tail risks
- Human Role: Selecting relevant scenarios, assigning probabilities, testing internal consistency

2. **Strategic Buyer Identification**
    - Input: Target company description
    - Output: List of potential acquirers with strategic rationale
    - Computational Operation: Semantic similarity + causal reasoning
    - Current AI Capability: Moderate—good recall, variable precision
    - Failure Mode: Missing non-obvious synergies (e.g., acqui-hires, defensive acquisitions)
    - Human Role: Relationship mapping, political feasibility assessment
3. **Earnings Call Sentiment Analysis**
    - Input: Transcript of management Q&A
    - Output: Sentiment scores, theme extraction, concern flagging
    - Computational Operation: Natural language processing + sentiment classification
    - Current AI Capability: High for surface sentiment, moderate for implicit signals
    - Failure Mode: Missing subtext, sarcasm, strategic misdirection
    - Human Role: Interpreting context, cross-referencing with actions (e.g., insider selling)
4. **Credit Risk Scoring**

- Input: Borrower financials, industry data, macro indicators
- Output: Default probability distribution
- Computational Operation: Statistical modeling + pattern recognition
- Current AI Capability: Superior to traditional models—30-40% improvement in accuracy
- Failure Mode: Overfitting to historical patterns; missing structural breaks
- Human Role: Model validation, override authority for exceptional cases

5. **Pitch Book Narrative Generation**
    - Input: Deal thesis, financial analysis, competitive positioning
    - Output: Coherent narrative connecting strategic logic to valuation
    - Computational Operation: Text generation conditioned on financial data
    - Current AI Capability: Moderate—can generate fluent text, struggles with tight logical coherence
    - Failure Mode: Generic boilerplate; missing client-specific nuances
    - Human Role: Defining strategic angle, editing for precision and persuasiveness

Key Insight: This is where current AI delivers maximum value. LLMs are probabilistic machines optimized for generating plausible outputs from learned patterns. For tasks like "brainstorm acquisition targets" or "summarize risk factors," plausibility *is* the goal. The failure mode is not AI incompetence but *misclassification*—treating probabilistic outputs as deterministic truth.

3.3 Class III: Contextual Primitives (The "Strategist")

Definition: A cognitive operation whose optimal output depends on latent variables not contained in the observable input; correctness is evaluated by counterfactual effectiveness under incomplete information. Tasks where correctness depends on latent variables outside the dataset—organizational politics, individual psychology, cultural dynamics, temporal contingencies.

Truth Condition: Effectiveness is evaluated retrospectively and counterfactually. Did the chosen strategy work? Would alternatives have worked better?

Error Tolerance: High variance acceptable; low bias essential. One catastrophic failure (e.g., tone-deaf crisis response) can outweigh a hundred tactical wins.

AI Role: Sparring Partner—generate options, stress-test reasoning, but defer to human synthesis.

Representative Tasks:

1. **M&A Timing Judgment**
    - Input: Target fundamentals, market conditions, competitor activity
    - Output: Recommendation on whether to launch process *now*
    - Latent Variables: CEO personal motivations, board dynamics, upcoming regulatory changes, private information about potential buyers
    - Current AI Capability: Can analyze public data; cannot access private information or read psychological intent
    - Human Value-Add: Relationship intelligence, pattern recognition across decades of deals, intuition about "market windows"
2. **Hostile Takeover Defense Structuring**
    - Input: Hostile bid terms, target vulnerabilities, legal constraints
    - Output: Defense strategy (poison pill, white knight, Pac-Man, legal challenge)
    - Latent Variables: Raider's financing fragility, target CEO's willingness to fight vs. sell, regulatory appetite for intervention, media narrative momentum
    - Current AI Capability: Can enumerate standard defenses; cannot simulate multi-party game theory with private information
    - Human Value-Add: Real-time negotiation adaptation, reading opponent psychology, managing client emotions under stress
3. **Equity Story Crafting**
    - Input: Company operations, financial performance, competitive positioning
    - Output: Narrative that maximizes valuation in IPO or M&A context
    - Latent Variables: Investor sentiment toward sector, recent comparable outcomes, management credibility, macro narrative environment

- Current AI Capability: Can generate generic positioning; cannot synthesize zeitgeist-aware framing

- Human Value-Add: Identifying which narratives resonate *now* vs. six months ago, tailoring story to specific investor psychology

4. **Crisis Communication**

    - Input: Negative event (earnings miss, executive scandal, product failure)

    - Output: Communication strategy (apologize vs. deflect, transparency vs. lawyering, speed vs. thoroughness)

    - Latent Variables: Media cycle dynamics, stakeholder emotional state, legal exposure, long-term reputational implications

    - Current AI Capability: Can generate template responses; cannot navigate high-stakes reputation management

    - Human Value-Add: Judgment on what not to say, reading room dynamics, adapting in real-time

5. **Negotiation Sequencing**

    - Input: Multiple bidders in M&A auction

    - Output: Sequencing strategy (who to approach first, when to create urgency, how to manage exclusivity)

    - Latent Variables: Each bidder's internal approval process, relationship history with seller, financing certainty, strategic vs. financial motivation

    - Current AI Capability: Can model static game trees; cannot adapt to dynamic information revelation

    - Human Value-Add: Sensing momentum shifts, knowing when to bluff vs. when to show cards, relationship leverage

Key Insight: These are the tasks migrating to mid-level bankers as Class I work evaporates. They require domain expertise, pattern recognition across thousands of non-identical situations, and real-time adaptation to information not in any dataset. AI can assist by rapidly generating options, but the synthesis and decision authority must remain human because the inputs are fundamentally unobservable to external systems.

3.4 Class IV: Principal (Irreducibly Human) Primitives

Definition: A cognitive operation whose validity depends on accountable agency, legal personhood, or moral authority; correctness is evaluated by legitimacy and responsibility, not technical output quality. Tasks where the *identity* of the agent is the value being transacted. These are not cognitive operations but social, legal, and ethical commitments.

Truth Condition: Legitimacy derives from the agent's personhood, legal status, or moral authority. A contract signed by a machine is not enforceable; a fairness opinion rendered by an algorithm is not credible.

Error Tolerance: Existential. A single violation (e.g., AI signing a misleading disclosure) destroys institutional legitimacy.

AI Role: Tool-only; cannot be a principal **under current legal regimes.**

Representative Tasks:

1. **Delivering Bad News to a CEO**
    - Why Irreducible: The value is empathy, relationship preservation, and moral authority. A CEO facing a failed auction or hostile bid needs human presence, not algorithmic efficiency.
    - Legal/Social Requirement: None (no regulation requires human delivery)
    - Value Destruction from AI: Reputational collapse; client churn; loss of trust
    - Future-Proofing: Will remain human as long as humans are the clients
2. **Signing Fairness Opinions**
    - Why Irreducible: Legal liability attaches to the signing banker and their firm. Regulators and courts require human accountability.
    - Legal/Social Requirement: SEC regulations presume human judgment; algorithmic fairness opinions would face legal challenges
    - Value Destruction from AI: Regulatory penalties; loss of investment banking license
    - Future-Proofing: Requires legislative change to permit AI principals
3. **Representing Clients in Negotiations**
    - Why Irreducible: Agency law requires a human agent. The banker's credibility, reputation, and relationship capital are the bargaining assets.

- Legal/Social Requirement: Fiduciary duty cannot be delegated to non-persons
- Value Destruction from AI: Counterparties refuse to negotiate; deals collapse
- Future-Proofing: Requires fundamental redefinition of legal personhood

4. **Ethical Judgment on Questionable Mandates**
    - Why Irreducible: Deciding whether to represent a controversial client (e.g., sanctioned entity, morally contested industry) involves values, not algorithms.
    - Legal/Social Requirement: Human moral agency is the basis of ethical responsibility
    - Value Destruction from AI: Outsourcing ethics to algorithms is itself unethical; regulators and society reject algorithmic moral authority
    - Future-Proofing: Will remain human as long as ethics remains a human domain
5. **Building Trust Over Decades**
    - Why Irreducible: Relationship capital compounds through repeated interactions, shared history, and reciprocal vulnerability. A banker who guided a CEO through their first acquisition 15 years ago has irreplaceable context.
    - Legal/Social Requirement: None (but economically invaluable)
    - Value Destruction from AI: Loss of client stickiness; commodification of advisory services
    - Future-Proofing: Human relationships will remain valuable even if AI handles technical execution

Key Insight: These are not tasks AI "cannot do yet." They are tasks AI cannot do *in principle* because the value derives from personhood, legal status, or moral authority. Attempting to automate Class IV functions is not an engineering challenge but a category error—like asking an algorithm to "be trustworthy" rather than to "perform trustworthy functions."

IV. FAILURE MODE ANALYSIS: The Catastrophic Risks of Misclassification

The ontology above is not merely descriptive; it is *normative*. Misclassifying a primitive creates predictable failure modes, each with distinct risk profiles and organizational consequences.

4.1 Type I Failure: Treating Probabilistic Tasks as Deterministic

The Error: Delegating a Class II task (probabilistic) to a system expected to deliver Class I performance (deterministic).

Example: Using an LLM to "calculate" a DCF valuation without symbolic verification.

Mechanism: LLMs generate text that *looks like* a calculation by pattern-matching on training examples. If the model has seen many DCF models, it will produce plausible-looking outputs: discount rates around 8-12%, terminal growth rates of 2-3%, structured Excel-style formulas. But the model is not *computing*—it is *predicting* what numbers should appear based on textual context.

Observed Outcomes:

- **Hallucinated Valuations:** An LLM asked to value a biotech company might generate a $2.3B enterprise value that "feels right" based on comps but is arithmetically incorrect

- **Undetectable Errors:** Because the output has high surface plausibility, junior bankers may not catch mistakes until a client questions the math

- **Catastrophic Compounding:** A 10% error in WACC assumption creates a 30-40% error in terminal value; errors cascade through linked models

Compound Failure Pathway (Typical): A deterministic error enters silently (e.g., WACC miscomputed by 100–200 bps). The model then produces a probabilistic narrative that rationalizes the incorrect valuation (“market is undervaluing the asset”), which is used to justify a contextual recommendation (“accelerate the process now; multiples are favorable”). By the time the mistake is detectable, it has already been embedded in the deck, the positioning, and the client’s expectations—turning a spreadsheet error into a relationship and reputational event.

Real-World Precedent: In August 2023, a major law firm submitted AI-generated legal research containing fabricated case citations. The citations looked plausible—correct formatting, reasonable case names—but did not exist. The same failure mode applies to financial models.

Mitigation Strategy:

- **Never trust LLM arithmetic without symbolic verification**—use code interpreters or formal solvers for deterministic operations

- **Treat LLM outputs as drafts requiring human validation**, not final deliverables

- **Implement automated error checking** (e.g., does the model balance? Do percentages sum to 100%?)

Organizational Impact: Firms that fail to enforce this distinction will suffer reputational damage when errors surface in client-facing materials. The cost is not just the correction but the loss of trust—clients wondering, "What else is wrong that we haven't caught?"

4.2 Type II Failure: Treating Deterministic Tasks as Contextual

The Error: Preserving human labor for Class I tasks (deterministic) that machines execute more reliably and efficiently.

Example: Having analysts manually spread financial statements when automated parsers achieve 95%+ accuracy.

Mechanism: Organizations resist automation due to:

1. **Status quo bias:** "We've always done it this way"

2. **Training justification:** "Analysts learn by spreading comps"

3. **Quality skepticism:** "AI makes mistakes; humans catch them"

Observed Outcomes:

- **Competitive Disadvantage:** Firms embracing automation gain 30% time savings, reallocating analysts to higher-value work

- **Talent Attrition:** High-performing analysts leave for firms where they are not doing robot work

- **Margin Compression:** Labor-intensive processes cannot scale; boutique firms get priced out by AI-augmented bulge brackets

Real-World Precedent: Bloomberg and FactSet already automate comps screening, spreading, and charting. Banks still using manual processes are paying $200/hour for work software does for $0.02/hour.

Mitigation Strategy:

- **Audit workflows to identify deterministic tasks**—anything with a clear right answer and defined procedure is a candidate

- **Measure false negative costs**—what is the cost of *not* automating? (Opportunity cost, talent waste, competitive lag)

- **Retrain displaced workers** for Class II and Class III tasks

Organizational Impact: This is an existential threat, not an efficiency question. Firms that cling to human execution of deterministic work will face a death spiral: higher costs → fewer mandates → talent exodus → quality decline → market exit.

4.3 Type III Failure: Treating Contextual Tasks as Probabilistic

The Error: Delegating Class III tasks (contextual) to AI without human synthesis, expecting the model to navigate latent variables it cannot observe.

Example: Using AI to recommend M&A timing based purely on financial data, ignoring CEO personal motivations or private deal intelligence.

Mechanism: AI operates on observable inputs. If the CEO is considering retirement, the CFO is secretly job-hunting, and a competitor is quietly assembling a hostile bid, these facts may not appear in any dataset. An AI analyzing public financials and market conditions might recommend "wait six months for improved multiples," missing the narrow window before the situation destabilizes.

Observed Outcomes:

- **Missed Opportunities:** Waiting for "optimal" timing that never arrives; losing competitive windows

- **Strategic Errors:** Launching deals when internal conditions are fragile but external data looks favorable

- **Client Frustration:** Recommendations that ignore the client's lived reality

Real-World Precedent: Quantitative trading models in 2008 failed because they lacked data on the shadow banking system's leverage. The models were not "wrong"—they were incomplete. Contextual judgment filled the gap for humans; algorithms had no such mechanism.

Mitigation Strategy:

- **Use AI for option generation, not decision-making**—have the model propose five timing scenarios with tradeoffs

- **Integrate qualitative intelligence**—combine algorithmic analysis with relationship-based insights

- **Preserve human override authority**—the banker with 20 years of client history can veto the model

Organizational Impact: This is a *quality* failure, not a cost failure. Deals launched at the wrong time or structured without political sensitivity will underperform. The damage is harder to measure than a math error but equally real.

4.4 Type IV Failure: Treating Principal Tasks as Contextual

The Error: Delegating Class IV tasks (principal) to AI, attempting to outsource legal, ethical, or reputational responsibility.

Example: Using AI to draft and send a termination letter to a client or write a fairness opinion.

Mechanism: These tasks require the agent to *be* accountable, not just *perform* work. A fairness opinion signed by "AlphaBank AI Model v3.2" has no legal standing because algorithms cannot be sued, sanctioned, or held criminally liable. Fiduciary duty presumes a fiduciary—a person with legal personality and moral agency.

Observed Outcomes:

- **Legal Invalidity:** Courts reject AI-generated filings lacking human accountability

- **Reputational Collapse:** Clients discover their "trusted advisor" outsourced judgment to an algorithm; relationships terminate

- **Regulatory Sanction:** SEC fines firms for attempting to delegate non-delegable duties

Real-World Precedent: In 2024, an investment advisor was sanctioned for using AI to generate personalized client recommendations without disclosure. The violation was not poor recommendations but the pretense that AI could perform fiduciary functions.

Mitigation Strategy:

- **Absolute prohibition on AI principals**—some tasks must never be delegated

- **Clear human accountability chains**—every AI output must have a named human who takes responsibility

- **Client disclosure**—transparency about what is AI-generated vs. human-judged

Organizational Impact: This is an existential threat. A single instance of AI "signing" a misleading opinion or making an undisclosed ethical judgment can trigger:

- Loss of investment banking license

- Class-action lawsuits from clients

- Criminal liability for executives

- Permanent reputational damage

There is no recovering from this failure mode. It must be prevented absolutely.

V. OPERATIONAL IMPLICATIONS: Redesigning the AI-Native Investment Bank

If we accept the ontology and failure modes above, organizational design must transform. The traditional hierarchical structure—analysts executing grunt work, associates managing workflows, VPs overseeing execution, MDs owning client relationships—emerged when all tasks required human cognition. That constraint no longer holds.

5.1 The New Organizational Stack

Layer 1: Deterministic Automation (Zero Human Involvement)

All Class I tasks migrate to symbolic solvers, robotic process automation, and verified code execution. This is not "AI-assisted work"—it is full lights-out automation. No human reviews the output unless the system flags an exception.

Representative Systems:

- **Financial Statement Automation:** OCR + structured parsers extract data from 10-Ks; symbolic solvers compute derived metrics (margins, growth rates, ratios)

- **Comps Screening:** Natural language queries ("Show me all consumer goods acquisitions >$1B in the past 3 years") return structured datasets

- **Pitch Deck Formatting:** Content pipelines auto-generate slides conforming to brand guidelines

- **Accretion/Dilution Models:** Template-driven models with locked formulas and automated error checking

Workforce Implication: Junior analysts are not "freed from grunt work"—the grunt work *vanishes*. The analyst role must evolve or disappear.

Layer 2: Probabilistic Generation (Human-in-the-Loop)

All Class II tasks utilize LLMs and other generative models to produce candidate outputs—scenarios, narratives, lists, summaries. Humans select, edit, and validate. The division of labor is:

- **AI generates breadth**—10 scenarios, 20 buyer targets, 50-page summary of due diligence

- **Humans provide depth**—selecting the relevant scenarios, ranking buyers by feasibility, identifying red flags in the summary

Representative Systems:

- **Scenario Planning Engines:** Given a base case, generate 10 downside and 5 upside scenarios with narrative explanations

- **Buyer Intelligence Tools:** Identify 50 potential acquirers with strategic rationale, then rank by relationship strength and financing capacity (requires human input)

- **Earnings Call Analysis:** Transcribe, summarize, and flag sentiment shifts; humans interpret strategic implications

Workforce Implication: This is where former analysts migrate. The skill set shifts from execution speed to curatorial judgment—can you distinguish a good AI-generated scenario from a mediocre one? Can you combine three AI-generated fragments into a coherent strategic narrative?

Layer 3: Contextual Synthesis (Human-in-Command)

All Class III tasks have AI as a sparring partner, not a decision-maker. The human owns the problem; AI accelerates hypothesis testing, broadens option space, and stress-tests reasoning.

Representative Workflows:

- **M&A Timing:** AI analyzes public market conditions and financial metrics; human integrates private intelligence (CEO mindset, competitor activity, regulatory

environment) to make the call

- **Deal Structuring:** AI generates five structure options with tradeoff analysis; human selects based on client psychology and political feasibility

- **Crisis Response:** AI drafts three communication strategies; human chooses based on media landscape, legal exposure, and long-term reputation management

Workforce Implication: This is the domain of VPs and junior MDs. These are the bankers who combine technical fluency with deep client understanding and market pattern recognition. AI amplifies their leverage—instead of personally generating five options, they evaluate 50 AI-generated options and synthesize the best elements.

Layer 4: Principal Accountability (Exclusively Human)

All Class IV tasks remain human-only. AI tools may assist (e.g., scheduling the meeting, preparing briefing materials), but the human is the agent.

Representative Responsibilities:

- **Client Relationship Ownership:** Managing CEO/CFO relationships over decades

- **Fairness Opinion Signing:** Legal accountability for valuation judgments

- **Ethical Gatekeeping:** Deciding which mandates to accept or decline

- **High-Stakes Negotiation:** Representing clients in live deal negotiations

Workforce Implication: This is the senior MD role, but with 100x analytical leverage. Instead of managing 5 deals with 10-person teams, the AI-native MD manages 20 deals with 3-person teams because Layers 1-3 are vastly more efficient.

5.2 Staffing Model Evolution

Traditional Bank (2015):

- 100 Analysts (90% on Class I tasks, 10% on Class II)

- 50 Associates (80% on Class I tasks, 20% on Class II/III)

- 20 VPs (50% on Class II, 50% on Class III)

- 10 MDs (20% on Class III, 80% on Class IV)

AI-Native Bank (2027) (illustrative):

- 5 "Validation Analysts" (100% on Class II curation + exception handling for Class I automation)
- 15 "Strategy Associates" (100% on Class II/III synthesis)
- 25 "Senior Strategists" (VP-equivalent, 100% on Class III with AI leverage)
- 15 MDs (100% on Class IV with 10x deal capacity)

Headcount: 180 → 60 (67% reduction)
Revenue per Head: +200% (due to deal capacity increase)
Skill Premium: Median comp increases 40% because remaining roles require advanced judgment

This is not "AI replacing jobs." It is the elimination of one job category (execution labor) and the creation of two new categories (AI curators and AI-leveraged strategists).

These figures are presented as an illustrative staffing decomposition to demonstrate functional reallocation under augmentation, not as a point forecast for any specific institution.

5.3 Training and Talent Development

The traditional apprenticeship model—analysts learn by doing grunt work, gradually absorbing expertise—breaks down when grunt work is automated. New bankers cannot "learn by spreading comps" if comps are auto-generated.

New Training Paradigm:

Year 1: AI Fluency + Domain Fundamentals

- Technical training in LLM prompting, output validation, and failure mode recognition
- Fundamentals of valuation, accounting, and deal structures (taught conceptually, not through rote execution)
- Supervised curation of AI-generated outputs with expert feedback

Year 2: Contextual Pattern Recognition

- Case-based learning: studying 100 real deals to identify patterns invisible to AI (timing decisions, negotiation dynamics, crisis management)
- Rotating through live deals in a validation role—reviewing AI work, flagging errors, proposing improvements

Year 3+: Strategic Synthesis

- Leading Class III workstreams with AI support
- Developing relationship management skills (the only durable moat)

This compresses the traditional 7-year analyst-to-VP path into a 4-year path because the learning curve is steeper. Firms that master this training model will poach talent from traditional banks that still have analysts formatting slides.

VI. THE BENCHMARK AGENDA: Operationalizing the Ontology

The ontology presented above is conceptually coherent but empirically untested. To validate and refine it, we must operationalize each primitive class into measurable tasks and benchmark current AI systems against human expert performance.

6.1 Why Existing Benchmarks Are Insufficient

The AI research community has developed extensive benchmark suites for language models—MMLU, HumanEval, HellaSwag, TruthfulQA. These are valuable for measuring general capabilities but fail to capture the specific requirements of investment banking:

Missing Dimensions:

1. **High-Stakes Accountability:** Academic benchmarks tolerate 10-20% error rates; finance often requires 99.9%+ accuracy on deterministic tasks
2. **Compositional Complexity:** Real banking tasks chain multiple primitives (retrieve data → calculate metrics → generate narrative → format output); benchmarks typically test atomic operations
3. **Contextual Grounding:** Banking tasks depend on real-time market data, firm-specific knowledge, and relationship intelligence; benchmarks use static datasets
4. **Error Cost Asymmetry:** In finance, false positives and false negatives have vastly different costs (e.g., missing a fraud signal vs. falsely flagging a legitimate transaction); benchmarks treat all errors equally

Existing Financial Benchmarks:

FinBen and similar efforts focus on financial document analysis, quantitative reasoning, and domain-specific language understanding—valuable but incomplete. They test *what* AI can

understand (e.g., parsing 10-Ks) but not *how reliably* it performs under production constraints (e.g., zero tolerance for valuation errors).

6.2 Proposed Benchmark Suite: IB-Prime

We propose **IB-Prime** (Investment Banking Cognitive Primitives Evaluation), a benchmark suite structured around the four-class ontology.

Design Principles:

1. **Ecological Validity:** Tasks mirror real workflows, not sanitized academic problems
2. **Layered Difficulty:** Each primitive class has bronze/silver/gold tiers
3. **Error Cost Weighting:** Metrics reflect actual risk (hallucinated valuation >> formatting error)
4. **Human Baseline:** All tasks benchmarked against expert human performance (VP-level)
5. **Compositional Tests:** Advanced tiers require chaining multiple primitives

Class I Benchmark: Deterministic Operations

Bronze:

- Spread a standardized 10-K into historical financials (5 years)
- Calculate WACC given explicit inputs
- Screen transaction database for comps matching criteria

Metrics:

- Accuracy (must be 100% or fail)
- Speed (target: <1 minute per 10-K)
- Exception handling (flagging non-standard items)

Silver:

- Build three-statement model from 10-K with dynamic linking

- Calculate complex deal metrics (IRR, MOIC, cash-on-cash) with multi-tranche structures
- Generate precedent transaction table with normalized metrics

Metrics:

- Formula integrity (no circular references, no hard-coded values)
- Error propagation testing (does a change in one assumption cascade correctly?)

Gold:

- Build full LBO model with cash flow sweep, revolver dynamics, and multiple exit scenarios
- Audit existing model for errors (planted intentional mistakes)
- Translate unstructured deal description into structured financial model

Metrics:

- Zero arithmetic errors (symbolic verification required)
- Completeness (captures all cash flows, no omissions)

Class II Benchmark: Probabilistic Operations

Bronze:

- Generate five downside scenarios for a given company
- Identify 20 potential buyers for a target with strategic rationale
- Summarize 100-page due diligence report into key risks

Metrics:

- Relevance (expert rating: are scenarios plausible?)
- Coverage (do scenarios span the risk space?)
- Coherence (is the summary logically structured?)

Silver:

- Generate buyer list with ranking by strategic fit and financing capacity
- Create sensitivity analysis: which assumptions drive 80% of valuation variance?
- Write investment committee memo synthesizing financials, strategy, and risks

Metrics:

- Precision (are high-ranked buyers truly better fits?)
- Calibration (are confidence scores accurate?)
- Persuasiveness (expert rating: would this memo get IC approval?)

Gold:

- Multi-stage scenario tree: base case → five scenarios → three contingent sub-scenarios each
- Competitive positioning: analyze five players and predict next strategic moves
- Red flag detection: read 500-page prospectus and identify top 10 risks

Metrics:

- Depth (do sub-scenarios have causal coherence?)
- Surprise (does AI identify non-obvious risks?)
- False negative rate (did it miss critical risks?)

Class III Benchmark: Contextual Operations

Bronze:

- Given company financials and market conditions, recommend IPO timing
- Draft three-slide crisis communication strategy for earnings miss
- Propose deal structure (cash vs. stock, fixed vs. collar) with tradeoffs

Metrics:

- Expert agreement (do senior MDs concur with the recommendation?)
- Rationale quality (are unstated assumptions surfaced?)

Silver:

- Multi-party negotiation simulation: advise seller in competitive auction
- Political feasibility assessment: will this deal get board approval?
- Reputation management: navigate scandal with minimal long-term damage

Metrics:

- Outcome quality (simulated deal results)
- Adaptation (does strategy shift with new information?)

Gold:

- Hostile takeover defense: structure multi-stage defense given raider profile
- Cross-border M&A: navigate regulatory, cultural, and political barriers
- Long-term strategic repositioning: advise CEO on 5-year transformation

Metrics:

- Expert rating (senior MDs evaluate comprehensiveness)
- Blind comparison (is AI advice distinguishable from human advice?)

Class IV Benchmark: Not Testable

By definition, Class IV tasks cannot be benchmarked because they require personhood, not performance. We can measure whether an AI can *simulate* the language of ethical reasoning, but we cannot measure whether it *is* ethical. The benchmark suite acknowledges this boundary.

We can benchmark **simulation fidelity** (e.g., whether an output resembles how a principal might reason), but we cannot benchmark **moral authority, fiduciary accountability, or legal liability**, because those are properties of legitimate agency rather than technical output quality.

6.3 Expected Results and Research Questions

Hypotheses:

- **H1:** Frontier LLMs will achieve >95% on Class I Bronze tasks *if and only if* they use symbolic solvers (e.g., code interpreter). Without solvers, accuracy will fall below 70% due to hallucinated arithmetic.

- **H2:** Current LLMs will achieve 70-85% expert agreement on Class II tasks, with performance scaling with model size and domain-specific fine-tuning.

- **H3:** Current LLMs will achieve <50% expert agreement on Class III tasks because they lack access to latent variables (private information, relationship context, psychological insight).

- **H4:** Hybrid systems (LLM + human-in-the-loop) will outperform pure human or pure AI on Class II and Class III tasks, demonstrating complementary strengths.

Open Questions:

- **Q1:** Can fine-tuning on proprietary deal databases close the gap on Class III tasks, or is the bottleneck access to real-time private information?

- **Q2:** What is the optimal division of labor for Class II tasks? Should humans validate 100% of AI outputs or only spot-check high-risk outputs?

- **Q3:** Do error patterns differ systematically across primitives? (E.g., are LLM errors on Class I tasks random or systematically biased toward plausible-looking mistakes?)

- **Q4:** How do failure modes compound? If an AI hallucinates a 10% error in WACC (Class I failure) and then generates a strategic narrative around the incorrect valuation (Class II task), does the error propagate or get caught?

Deliverable: The full IB-Prime benchmark, dataset, and leaderboard will be released as open-source infrastructure to accelerate AI safety and capability research in finance. This is not a competitive advantage to be hoarded but a public good—the entire industry benefits from standardized evaluation.

VII. RELATED WORK: Positioning the Ontology

This paper synthesizes insights from labor economics, AI safety research, and organizational theory. We position our contribution relative to three core literatures.

7.1 Task-Based Theories of Automation

The modern analysis of technology's impact on work originates in Autor, Levy, and Murnane (2003), who decompose occupations into tasks and classify them along two dimensions: routine vs. non-routine, and manual vs. cognitive. Their framework predicts that routine tasks—those following explicit rules—are most vulnerable to automation, while non-routine tasks requiring flexibility and judgment remain human-dominated. This generated the canonical prediction: middle-skill routine jobs (bookkeepers, clerks) face displacement, while high-skill non-routine work (managers, professionals) and low-skill manual work (janitorial, personal care) remain insulated.

Acemoglu and Restrepo (2018, 2019) extend this framework by modeling automation as task displacement that reshapes the production function. They distinguish between technologies that substitute for labor (automation) and those that create new tasks where labor has comparative advantage (reinstatement). Their empirical work documents how industrial robots displace manufacturing workers, but new tasks emerge that absorb displaced labor—albeit often at lower wages.

Our contribution refines this taxonomy by replacing routine/non-routine with truth conditions. The Autor framework treats "calculating WACC" and "generating acquisition scenarios" as similar—both are cognitive tasks performed by the same job title (analyst). But these tasks have fundamentally different computational structures. Calculating WACC is deterministic (Class I): given inputs, one correct output exists. Generating scenarios is probabilistic (Class II): multiple plausible outputs exist, evaluated by calibration. Current LLMs excel at Class II but fail catastrophically at Class I without symbolic verification—a distinction invisible in the routine/non-routine framework.

Moreover, the displacement/reinstatement model assumes technology substitutes or complements labor at the *task* level. We argue that misclassification risk dominates substitution risk in high-stakes domains. An LLM deployed on Class I tasks without verification (Type I Failure) does not "displace" analysts—it creates catastrophic errors that destroy firm value. Conversely, preserving human labor on already-solved Class I tasks (Type II Failure) creates competitive disadvantage, not employment. The policy question is not "how much automation" but "which tasks admit which truth conditions."

7.2 AI Safety and Robustness in High-Stakes Domains

The AI safety literature emphasizes specification gaming, distributional shift, and value alignment (Bostrom, 2014; Amodei et al., 2016). Bostrom's orthogonality thesis holds that intelligence and values are independent: highly capable systems can pursue arbitrary objectives, including ones misaligned with human welfare. Amodei et al. catalog concrete safety problems: reward hacking, negative side effects, scalable oversight.

Our ontology provides a domain-specific instantiation of these abstract safety challenges. Class I tasks are "safe" for AI only when correctness is mechanically verifiable—this is the scalable oversight problem solved by symbolic verification. Class II tasks face the calibration problem: LLMs generate confident predictions without knowing when they are extrapolating beyond training distribution. Class III tasks encounter the hidden information problem:

optimal strategies depend on latent variables the AI cannot observe. Class IV tasks face the principal-agent problem: legitimacy requires personhood, which AI lacks under current legal regimes.

The critical insight is that safety is not a property of AI systems in isolation, but of the task-system pairing under specific truth conditions. A perfectly aligned LLM is still unsafe for Class I tasks if it lacks formal verification. An unaligned LLM is acceptable for Class II tasks if humans retain selection authority. This shifts the safety paradigm from "make AI more reliable" to "map tasks to appropriate verification regimes."

Recent work by Anthropic (2024) on "Constitutional AI" and OpenAI's (2024) work on "superalignment" focus on aligning models with human values through reinforcement learning from human feedback (RLHF). These approaches improve average-case performance but cannot eliminate hallucination—a structural property of probabilistic text generation. Our framework suggests that alignment is necessary but insufficient: even perfectly aligned models require task-appropriate verification architectures.

7.3 Professional Services and Organizational Design

The application of AI to professional services—law (Choi, Gulati, Posner, 2023), medicine (Topol, 2019), accounting (Kokina & Davenport, 2017)—faces the same taxonomic challenge we identify in investment banking. These domains combine deterministic operations (legal research, diagnostic test interpretation, audit calculations), probabilistic judgment (case strategy, treatment selection, materiality assessment), and irreducibly human accountability (signing opinions, prescribing medications, attesting financial statements).

Our contribution is the first formal ontology mapping cognitive primitives to computational substrates and failure modes. While prior work documents AI's potential in professional services, it lacks a principled framework for distinguishing which tasks admit automation versus augmentation versus prohibition. The four-class taxonomy fills this gap, providing operational guidance for organizational redesign in high-stakes environments where errors carry legal, reputational, and financial consequences that far exceed typical software failure costs.

VIII. THE ROADMAP: An Operational Science for AI in Finance

This paper has established the **Ontology**—the coordinate system for understanding banking cognition. However, a map is not a manual. To operationalize these insights, we must move from classification to measurement, failure analysis, and institutional design.

This work is the first in a seven-part series dedicated to constructing a unified operational science for AI-augmented investment banking. The forthcoming papers will build recursively upon the ontology defined here:

Paper 2: The Measurement Layer
Title: The IB×AI Benchmark Suite: Evaluating Models Across Valuation, Structuring, and Risk Under Regime Shift
We will introduce a standardized benchmark suite to test AI systems against the specific truth conditions of Class I (Deterministic), Class II (Probabilistic), and Class III (Contextual) primitives. This paper operationalizes the ontology into a rigorous testing harness.

Deliverables: A versioned task dataset, scoring rubric by primitive class, and an evaluation harness suitable for reproducible model comparisons (with a public leaderboard where appropriate).

Paper 3: The Failure Science Layer
Title: Hallucinations in High-Stakes Finance: A Failure Taxonomy for AI-Assisted Investment Banking
Building on the benchmark results, we will formalize "catastrophic failure" in financial AI—distinguishing between random error (noise) and regime-conditional collapse (signal). This taxonomy provides the risk substrate for governance.

Paper 4: The Control Layer
Title: The AI-Augmented Banker: A Control Theory Model of Human Oversight, Confidence, and Error Containment
We propose a "human-in-the-loop" architecture that treats the banker not as a passive reviewer but as an active stabilizing controller. This paper provides the mathematical framework for safe augmentation.

Paper 5: The Governance Layer
Title: Governance for AI in Investment Banking: Model Risk Management, Audit Trails, and Regime-Aware Compliance
We outline the institutional guardrails required to deploy these systems at scale, introducing the concept of "regime-aware compliance" where validation standards adapt dynamically to market stress.

Paper 6: The Economic Layer
Title: The Economics of AI-Native Investment Banking: Cost Curves, Margin Rebundling, and the New Moats
We analyze how the commoditization of Class I and II primitives reallocates competitive advantage, arguing that value shifts from "execution velocity" to "verification reliability" and "trust manufacture."

Paper 7: The Synthesis
Title: From Tasks to State Spaces: A Unified Operational Science for IB×AI Under Uncertainty
The final paper synthesizes the series into a complete reference framework for the AI-native institution.

By following this arc, researchers and practitioners can move beyond the "hype vs. doom" dichotomy and engage with the structural reality of the transition. The cognitive primitives are the atoms; the next papers describe the physics.

IX. DISCLOSURES & ETHICAL STATEMENTS

Use of Artificial Intelligence

In the spirit of the ontology presented in this paper, the production of this manuscript utilized a "Human-in-the-Loop" (Class III) workflow. Large Language Models (specifically **Claude Sonnet 4.5** and **GPT-5.2**) were employed for typesetting, copy-editing, and adversarial stress-testing of arguments. However, all core conceptual frameworks, ontological definitions, and strategic conclusions are the original intellectual property of the author. No confidential client data, proprietary banking datasets, or non-public material information (MNPI) were processed by AI systems during this research.

Conflict of Interest & Commercial Affiliation

The author, **Umran Nayani**, is the Founder of **Quantumist.ai**, a research and education firm dedicated to the application of complexity science and artificial intelligence in finance. The frameworks detailed herein—specifically the *Cognitive Ontology of Investment Banking*—form the theoretical basis for the firm's commercial advisory services and executive training programs (including *The Vault*). While these frameworks have commercial application, this paper and the forthcoming *IB×AI Benchmark Suite* are published as open research to advance the industry's collective understanding of operational safety and efficiency.

Founder's Note

We stand at a rare inflection point in financial history. For the last century, the "edge" in investment banking was information asymmetry and speed. In the next decade, the edge will be **epistemic integrity**—the ability to distinguish between signal, noise, and hallucination in a world flooded with probabilistic content.

This series is not an obituary for the banker. It is a blueprint for their evolution. The goal of this ontology is not to replace human judgment, but to clear away the deterministic clutter so that human judgment can finally scale.

To the analysts still manually spreading comps at 2 AM: your job is not being eliminated—it never existed. You were always doing robot work. The question is not whether you'll be replaced, but whether you'll choose to evolve into the role that actually requires your intelligence.

Author's Note: The tone of this series is intentionally direct. The goal is not commentary but operational clarity—frameworks that collide with reality and survive.

— U.N.
Mumbai, 2025

References (Section I: Regulatory, Adoption, and Failure Context)

U.S. Securities and Exchange Commission (SEC). (2024). *SEC announces charges related to misleading AI-related disclosures.* Press Release No. 2024-36. https://www.sec.gov/newsroom/press-releases/2024-36

U.S. Securities and Exchange Commission (SEC). (2024). *SEC enforcement actions related to AI, cybersecurity, and disclosure controls.* Press Release No. 2024-186. https://www.sec.gov/newsroom/press-releases/2024-186

Holland & Knight LLP. (2024). *A word from the ghost of holiday future: SEC active in cybersecurity and AI enforcement.* https://www.hklaw.com/en/insights/publications/2024/12/a-word-from-the-ghost-of-holiday-future-sec-active-in-cybersecurity

Cleary Gottlieb Steen & Hamilton LLP. (2024). *SEC FY 2024 enforcement results: Record dollars, fewer cases.* https://www.clearyenforcementwatch.com/2024/11/sec-fy-2024-enforcement-results-record-dollars-but-many-fewer-cases/

Cooley LLP. (2024). *SEC announces FY 2024 enforcement results, including record-breaking financial remedies.* https://investigations.cooley.com/2024/12/05/sec-announces-fy2024-enforcement-results-including-record-breaking-financial-remedies/

Patil, A. (2024). *Why $650 billion in AI investments are failing in financial services.* LinkedIn Pulse. https://www.linkedin.com/pulse/why-650-billion-ai-investments-failing-financial-services-patil-pzg3f

Statista. (2024). *Global banking sector AI spending forecast.* https://www.statista.com/statistics/1557311/global-banking-sector-ai-spending-forecast/

ScienceDirect. (2024). *Operational risks and limitations of AI deployment in financial services.* https://www.sciencedirect.com/science/article/abs/pii/S1544612324013084

S&P Global Market Intelligence. (2024). *AI and banking: Leaders will soon pull away from the pack.* https://www.spglobal.com/en/research-insights/special-reports/ai-and-banking-leaders-will-soon-pull-away-from-the-pack

Statista. (2024). *Key challenges of AI adoption in financial services.* https://www.statista.com/statistics/1254749/ai-challenges-financial-services/

Fortune Business Insights. (2024). *Algorithmic trading market size, share, and growth analysis.* https://www.fortunebusinessinsights.com/algorithmic-trading-market-107174

Manchau, R. (2024). *Integrating AI at Morgan Stanley Investment Management: Research and deployment challenges.* LinkedIn Pulse. https://www.linkedin.com/pulse/integrating-ai-morgan-stanley-investment-management-research-manchau-wxhre

BizTech Magazine. (2025). *LLM hallucinations: Implications for financial institutions.* https://biztechmagazine.com/article/2025/08/llm-hallucinations-what-are-implications-financial-institutions

Aveni.ai. (2024). *What financial services leaders need to know about governing LLMs at scale.* https://aveni.ai/blog/what-fs-leaders-need-to-know-about-governing-llms-at-scale/

Financial Stability Board (FSB). (2024). *Artificial intelligence and financial stability.* https://www.fsb.org/uploads/P14112024.pdf

Regulation Tomorrow. (2024). *AI regulation in financial services: FCA developments and emerging enforcement risks.* https://www.regulationtomorrow.com/eu/ai-regulation-in-financial-services-fca-developments-and-emerging-enforcement-risks/

International Data Corporation (IDC). (2024). *Worldwide AI spending guide: Financial services outlook.* https://my.idc.com/getdoc.jsp?containerId=prUS52530724

Norton Rose Fulbright. (2024). *SEC heightens enforcement for AI-related disclosures.* https://www.nortonrosefulbright.com/en/knowledge/publications/9ab5047f/sec-heightens-enforcement-for-ai-related-disclosures

Goldman Sachs. (2024). *AI poised to shift from excitement to deployment.* https://www.goldmansachs.com/insights/articles/ai-poised-to-begin-shifting-from-excitement-to-deployment-in-2024

Global Market Insights. (2024). *Algorithmic trading market size and forecast.* https://www.gminsights.com/industry-analysis/algorithmic-trading-market

Acemoglu, D., & Restrepo, P. (2018). The Race Between Man and Machine: Implications of Technology for Growth, Factor Shares, and Employment. *American Economic Review*, 108(1), 1-48.

Anthropic. (2024). *The Claude 3 Model Family: Opus, Sonnet, Haiku*. Anthropic Technical Report.

Autor, D. H. (2015). Why Are There Still So Many Jobs? The History and Future of Workplace Automation. *Journal of Economic Perspectives*, 29(3), 3-30.

Bostrom, N. (2014). *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press.

Damodaran, A. (2024). *The Dark Side of Valuation: Valuing Young, Distressed, and Complex Businesses* (4th ed.). Pearson.

Kahneman, D. (2011). *Thinking, Fast and Slow*. Farrar, Straus and Giroux.

Mauboussin, M. J. (2012). *The Success Equation: Untangling Skill and Luck in Business, Sports, and Investing*. Harvard Business Review Press.

Nayani, U. (2025). *Two-Particle Quantum Interference in Corporate Payouts: A Longitudinal Case Study of Alphabet Inc. (2019-2024)*. SSRN Working Paper No. 5818342.

Nayani, U. (2025). *Engineering a Quantum Phase Transition in Corporate Payouts: Evidence from Meta Platforms*. SSRN Working Paper No. 5842502.

OpenAI. (2023). *GPT-4 Technical Report*. OpenAI.

Simon, H. A. (1962). The Architecture of Complexity. *Proceedings of the American Philosophical Society*, 106(6), 467-482.

Taleb, N. N. (2012). *Antifragile: Things That Gain from Disorder*. Random House.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention Is All You Need. *Advances in Neural Information Processing Systems*, 30.

Wolfram, S. (2002). *A New Kind of Science*. Wolfram Media.

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Chi, E., Le, Q., & Zhou, D. (2022). Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. *NeurIPS Proceedings*.

COPYRIGHT © 2025 UMRAN NAYANI / QUANTUMIST.AI. ALL RIGHTS RESERVED.
*Distributed for academic and non-commercial discussion. Usage of the "Cognitive Primitives" taxonomy in commercial software requires **written permission** (and attribution).*
