A(I)gents for Finance: Sci-Fi, reality, and the philosophy of delegated Intelligence

Riccardo Bovetti

December 5, 2025

1 Premise and purpose

Evaluating the impact that the emergence of A(I)gents could have in the Finance function requires a conceptual foundation that goes beyond enthusiasm, technological curiosity, or surface-level operational considerations. Finance is, by its very nature, a domain grounded in responsibility, verification, judgment, and intentionality. The function exists to steward resources, ensure accuracy, maintain transparency, and uphold accountability across the entire organization and for these peculiar reasons, any conversation about autonomous or semi-autonomous agents must begin with a rigorous clarification of what is being delegated, to whom (or to what), and under which philosophical, organizational, and epistemic conditions such delegation can be considered legitimate. The purpose of this paper is to try to articulate an "elastic" framework for understanding limitation and opportunities of delegated intelligence in Finance. It does not focus on tools, vendors, system architectures, or short-term technological trends, all of which shift rapidly and often generate more noise than insight. Instead, it aims to build the conceptual scaffolding needed to interpret A(I)gents as organizational actors situated within a system of rules, meanings, constraints, and responsibilities. Such scaffolding in my opinion is essential because Finance cannot afford to adopt technologies based solely on their technical sophistication or market momentum. Every tool, system, or agent introduced into the function must be subjected to a form of epistemic due diligence that ensures it can be governed, justified, and integrated without compromising the integrity that defines the function itself. This requires (at least) three interconnected layers of examination. The first layer concerns conceptual clarity, which involves defining what an agent actually is, what distinguishes it from traditional automation and from generic AI, and what structural components determine its behavior. Without this clarity, the term "agent" risks becoming an empty signifier, a marketing label applied indiscriminately to any system that exhibits a degree of computational sophistication. The second layer concerns philosophical grounding, which involves establishing the ontological, epistemological, and ethical assumptions that make delegation acceptable in a function whose legitimacy depends entirely on trust and accountability. Philosophy here is not ornamental but foundational, it provides the grammar through which we can articulate what it means to delegate judgment, to distribute responsibility, and to maintain coherence in a system where some actors are human and others are not. The third layer concerns managerial (and governance) structure, which involves identifying the mechanisms, guardrails, and organizational conditions under which agents can be introduced without compromising integrity, reliability, or compliance. This layer addresses the practical translation of principles into policies, of concepts into controls. In this sense, the paper seeks to develop not only a vocabulary but also a grammar of delegated intelligence, one that enables Finance leaders to interpret agents as part of a broader systemic transformation rather than as isolated technological artifacts. The goal is not to provide a checklist for implementation, but to establish the intellectual foundation upon which responsible, thoughtful, and sustainable adoption can be built. The objectives of this work can be summarized as follows:

- to clarify the meaning of "agents" in the specific context of the Finance function, distinguishing them from other forms of automation and intelligence; to distinguish the hype surrounding agents from their operational reality, identifying what is genuinely transformative and what remains speculative or aspirational;
- to define the philosophical and governance principles that make delegated autonomy acceptable in a domain where trust is paramount;
- to introduce the algorithm of success and failure as a formal model for understanding why so many technological initiatives fail despite technical viability;
- and to outline the evolution of competency shapes, from the vertical expertise of the I-shaped professional to the augmented profile of the PI-shaped professional, capable of engaging with intelligent systems as cognitive partners rather than opaque black boxes.

2 The compass of "Three": a structural and conceptual constant

The architecture of this paper is intentionally built around triads, recurring structures of three elements that organize concepts, frameworks, and arguments throughout the entire work. This is not(only) an aesthetic preference, nor a stylistic quirk adopted for rhetorical effect. It reflects a deeper epistemological need rooted in how human cognition processes complexity and how meaning itself emerges from relationships rather than from isolated entities. Complex realities, particularly those involving socio-technical systems like Finance, are most intelligible when framed through structures that allow differentiation without fragmentation, comparison without reduction, and synthesis without deletion. Three is the first number that generates structure in a meaningful sense. One represents unity, indivisible and absolute, a totality that cannot be decomposed or analyzed without ceasing to be what it is. Two introduces polarity, contrast, and tension, creating opposition but offering no resolution, no third term that mediates or synthesizes. With two, we have thesis and antithesis but no movement toward synthesis. Only three creates a system, a structure in which relationships become directional rather than oppositional, dynamic rather than binary, capable of expressing evolution, hierarchy, and complexity without collapsing into dualism or dissolving into chaos. In mathematics, three is the smallest number that can define a plane, creating dimensionality beyond the line. In logic, the triad enables syllogistic reasoning, the movement from premises to conclusion. In theology, the concept of the Trinity represents the paradox of unity in multiplicity, three persons constituting one essence. In organizational theory, triadic structures recur whenever a phenomenon requires balance across competing principles, such as efficiency, effectiveness, and adaptability, or strategy, governance, and execution, or risk, return, and control. The triad allows comparison without forcing equivalence, permits tension without demanding conflict, and enables synthesis without erasing the distinctness of individual elements. It provides enough richness to avoid oversimplification, which is the trap of binary thinking, yet maintains enough structure to avoid cognitive overload, which is the risk of excessively complex taxonomies. In cognitive science, research has demonstrated that the human mind processes patterns more easily when they appear in groups of three. A triad provides the minimal scaffold for meaning-making, the smallest configuration that allows the mind to perceive relationships, recognize patterns, and construct narratives. In the context of A(I)gents in Finance, triads serve a second, equally important function beyond cognitive accessibility. They impose a rhythm on the argument, a cadence that guides the reader through what is otherwise an overwhelming landscape of technologies, narratives, fears, and expectations. Each triad in this paper marks a conceptual threshold, an axis along which Finance must evolve as it transitions from deterministic tools designed for predictable tasks toward semi-autonomous systems capable of interpretation, judgment, and purposeful action. These transitions are not technical upgrades but paradigm shifts, and paradigm shifts require not only new tools but new ways of thinking, new vocabularies, new frameworks for understanding what is happening and why it matters. Triads, in this sense, are not only analytical devices but navigational instruments. They operate as a compass, orienting the reader through terrain that is still being mapped, where boundaries are unclear and certainties are few. They break the illusion that the debate about agents is a binary choice between human judgment and machine autonomy, revealing instead a continuum of interactions, complementarities, and negotiations. They demonstrate that complexity is not chaotic but structured, that transformation is not arbitrary but patterned, and that understanding emerges not from simplification but from recognizing the relationships between elements. The recurring use of triads throughout this paper creates a cognitive rhythm that mirrors how humans process complex systems and how organizations internalize paradigm shifts. It is not enough to present information; that information must be structured in a way that makes it absorbable, memorable, and actionable. Triads achieve this by creating patterns that the mind can grasp intuitively, even when the content itself is abstract or philosophically dense.

2.1 π as the infinite sequence of triads

To anchor this structural principle symbolically, the number π is introduced as an additional conceptual frame. Pi, the mathematical constant representing the ratio of a circle's circumference to its diameter, is infinite in its decimal expansion, non-repeating in its pattern, and irreducible to any simpler form. Its sequence contains every possible combination of digits, every imaginable pattern, yet never stabilizes, never resolves, never reaches a point of closure. Pi is both perfectly defined mathematically and infinitely complex in its expression. Using Pi as the frame for "the need for many threes" means positioning this discussion in a universe where complexity is structured but not linear, where phenomena are governed by principles but not predictable in their outcomes, where order coexists with inexhaustibility. Pi becomes a metaphor for the Finance function in the age of AI: a domain characterized by infinite data flows, endless combinations of variables, and continuous emergence of new patterns, yet ultimately constrained by finite decisions that must be made within bounded timeframes, under specific regulatory requirements, and in accordance with clear standards of accountability. Just as Pi's digits extend infinitely while still being contained within a precise mathematical relationship, so too does the challenge of agents in Finance extend across countless dimensions (technological, organizational, philosophical, ethical, regulatory) while still being bounded by the fundamental nature of the function itself. Finance cannot escape into pure experimentation or unstructured innovation; it must always return to its core mandate of stewardship, accuracy, and accountability. Pi reminds us that complexity does not mean chaos, that infinity does not mean arbitrariness, and that structure persists even in the midst of inexhaustible variation.

3 Why today: the structural conditions of the question

The relevance of agents in Finance emerges not from a single breakthrough or a singular moment of technological revelation, but because several structural forces, long in the making and individually insufficient, are finally converging in a configuration that makes their collective impact transformative. These forces have been developing over years, sometimes decades, each following its own trajectory, but only recently have they reached a threshold of maturity and scale where their interaction creates qualitatively new possibilities. Their convergence does not automatically generate intelligence, a crucial distinction often lost in popular discourse, but it radically alters the cost, speed, and scale at which intelligence-like behaviors can be orchestrated, simulated, and deployed within organizational systems. This convergence is neither accidental nor inevitable. It is the result of specific technological, economic, and societal developments that have collectively created an environment in which certain forms of computational action, previously confined to research labs or science fiction narratives, become operationally plausible and economically viable. Understanding why agents are possible today requires understanding not just the technologies themselves but the ecosystem conditions that enable their existence.

3.1 The first structural triad: the enablers

The first set of forces consists of three enablers that collectively remove the constraints that historically made agent-like systems impractical, unaffordable, or impossible to integrate into real-world organizational processes. The first enabler is data abundance, which represents far more than a mere increase in volume or velocity. It constitutes a fundamental transformation in the ontology of organizational information. Finance historically operated within an informational regime characterized by retrospective data (generated after events happened), curated datasets (carefully selected and validated before use), and periodic reporting cycles (monthly closes, quarterly reviews, annual audits). This regime was inherently backward-looking, focused on recording what had already happened rather than interpreting what was currently unfolding or anticipating what might occur next. Today, however, data is generated continuously, in real time, as a byproduct of nearly every transaction, interaction, and operational event. This data is not merely quantitative but also behavioral, contextual, relational, and unstructured. The result is a shift in the fundamental bottleneck of organizational intelligence: the constraint is no longer access to information but the capacity to extract meaning from the overwhelming flow of information. Agents become relevant precisely because they can operate within this new informational regime, processing streams of data that exceed human cognitive bandwidth and identifying patterns, anomalies, or opportunities that would remain invisible to traditional analytical methods. The second enabler is low-cost compute at scale, which breaks a constraint that has governed organizational decision-making for more than a century. Historically, computational capacity was a scarce, expensive, and carefully rationed resource. Organizations invested in mainframes, data centers, and dedicated infrastructure, and these investments required long planning cycles, significant capital commitments, and centralized control. The marginal cost of computation was high, which meant that experimentation was expensive, iteration was slow, and only high-value use cases could justify the investment. The emergence of cloud computing, distributed architectures, and elastic resource provisioning has fundamentally changed this calculus. Compute has become a commodity, available on demand, scalable up or down dynamically, and priced according to actual usage rather than projected capacity. The marginal cost of intelligence has collapsed, not because intelligence itself has become cheaper to produce (it remains computationally intensive), but because the infrastructure required to support it is no longer a fixed capital investment but a variable operational expense. This transformation enables rapid prototyping, continuous learning, and adaptive systems that evolve based on feedback rather than being designed once and deployed statically. The third enabler is interoperability and digital globalization, which together dissolve the organizational boundaries that once confined systems to isolated domains. In the past, enterprise systems were proprietary, monolithic, and incompatible, each operating according to its own data models, communication protocols, and application logic. Integration was difficult, expensive, and fragile, requiring custom middleware, manual reconciliation, and ongoing maintenance. Today, the proliferation of APIs (Application Programming Interfaces), standard data formats (JSON, XML, RDF), shared ontologies, and cloud-native architectures has created an environment in which systems can communicate fluidly, exchange information transparently, and coordinate actions across organizational and even geographical boundaries. This shift transforms Finance from a closed, ledger-keeping function operating within the walls of a single enterprise into a node embedded within a wider digital ecosystem that could include suppliers, customers, banks, regulators, tax authorities, auditors, and markets. Agents become possible not because they possess superhuman intelligence but because the environment itself is becoming machine-navigable, structured in ways that computational actors can interpret, traverse, and act upon with increasing autonomy. These three forces, data abundance, low-cost compute, and interoperability, enable agents not by creating cognition in any strong sense but by removing the frictions, economic constraints, and technical barriers that made cognition-like behavior operationally irrelevant in earlier eras. They establish the necessary conditions for agent systems but do not guarantee their success, which depends on how organizations design, govern, and integrate them.

3.2 The second structural triad: the distortions

At the same time that enablers create the conditions for agents, a second set of forces amplifies expectations and distorts perceptions, generating a gap between what is technically possible and what is believed to be imminent or inevitable. These distortions follow a pattern familiar from previous waves of technological enthusiasm, the pattern of the technology bubble, where narratives outpace realities, where investment flows toward potential rather than proven value, and where the rhetoric of transformation becomes detached from the constraints of implementation. The first distortion is media amplification, which operates through the accelerating velocity of information dissemination and the structural incentives of contemporary media ecosystems. Technologies are communicated faster than they can be understood, interpreted, or critically evaluated. News cycles favor novelty over nuance, breakthroughs over qualifications, and visionary claims over cautious assessments. The result is a cognitive asymmetry between reality and perception: what exists is conflated with what is promised, what works in controlled conditions is presented as ready for universal deployment, and what requires significant organizational adaptation is portrayed as plug-and-play. Media amplification does not create false information so much as it creates temporal distortion, collapsing the distance between early-stage research, prototype demonstrations, pilot deployments, and mature, scalable solutions. For Finance leaders navigating this landscape, the challenge is distinguishing signal from noise, substance from spectacle. The second distortion is investment overshoot, which occurs when capital flows into a domain not proportionally to its current maturity or demonstrated impact but proportionally to the narrative momentum and competitive FOMO (fear of missing out). Venture capital, private equity, and corporate Research and Development budgets flood into agent-related technologies, creating a market dynamic in which solutions are declared transformative, disruptive, or revolutionary before they have been proven reliable, before their failure modes are understood, before their operational costs are transparent, and before their regulatory implications are resolved. This overshoot creates perverse incentives: companies are rewarded for announcing ambitious roadmaps rather than delivering incremental value, for generating press coverage rather than operational results, for claiming AGI (Artificial General Intelligence) proximity rather than building robust, specialized tools. The consequence for adopters, particularly in risk-averse domains like Finance, is a marketplace saturated with immature products, exaggerated capabilities, and unclear value propositions. The third distortion is regulatory lack or incompleteness, which creates a vacuum of norms, standards, and legal frameworks capable of governing agent systems in domains where legitimacy and accountability are paramount. Regulation develops slowly, incrementally, and reactively, often lagging years or even decades behind technological change. In Finance, where the function's authority derives from its adherence to established rules, its ability to produce auditable records, and its alignment with societal expectations of fairness and transparency, this vacuum generates both enthusiasm and fear. Enthusiasm arises from the temporary freedom to experiment without constraint, to explore possibilities before they are foreclosed by regulation. Fear arises from the knowledge that this freedom is provisional, that regulatory intervention is inevitable, and that systems designed in the absence of clear rules may become liabilities when those rules eventually emerge. The result is a period of profound uncertainty in which organizations must decide whether to lead, adopting agents early and shaping best practices that may influence future regulation, or to wait, observing others and allowing the regulatory landscape to stabilize before committing resources. These three forces, media amplification, investment overshoot, and regulatory incompleteness, inflate the idea of agents beyond their actual current capabilities, creating an environment where speculation often precedes understanding, where expectations are set by marketing rather than by evidence, and where the distance between promise and reality becomes difficult to measure.

3.3 Why this is becoming (or already is) a bubble

A technological bubble forms when expectations grow faster than the foundations required to support those expectations. Bubbles are not characterized by the absence of real innovation but by the misalignment between the rate of narrative expansion, the rate of capital investment, and the rate of organizational absorption. In the current moment surrounding agents, all the conditions are present for replicating the well-known historical pattern of technological bubbles, and three specific misalignments are accelerating this trajectory. First, technological capability grows linearly, constrained by the pace of scientific research, engineering refinement, and empirical validation. Progress occurs through incremental improvements, careful testing, and gradual accumulation of knowledge. Second, narrative expectations grow exponentially, amplified by media cycles, reinforced by investment momentum, and propelled by competitive pressures. Each announcement of a new capability, each demonstration of a novel use case, and each visionary claim by a charismatic founder generates a wave of attention that exceeds the significance of the underlying development. Third, organizational readiness grows slowly and unevenly, constrained by institutional inertia, cultural resistance, legacy systems, regulatory compliance, and the simple fact that changing how people work is far more difficult than changing the tools they use. This mismatch creates structural tension. Agents are talked about as autonomous decision-makers capable of replacing human judgment in complex, high-stakes scenarios. Yet the organizational systems into which they must be integrated are still structured around deterministic workflows, sequences of predictable steps governed by explicit rules, formal controls, periodic reconciliations, and multi-layered approval chains. The architecture of Finance is designed to eliminate ambiguity, to enforce predictability, and to ensure traceability. Semi-autonomous actors, however, introduce a fundamentally different logic, one characterized by interpretation rather than execution, approximation rather than precision, and adaptive behavior rather than fixed rules. The bubble, therefore, is not primarily in the technology itself but in the projections placed upon it. The technology works within limits; the projections ignore those limits. The technology requires careful integration; the projections assume seamless adoption. The technology introduces new risks; the projections focus exclusively on new opportunities. When this misalignment becomes unsustainable, when the gap between expectation and reality becomes undeniable, a correction occurs. Valuations fall, enthusiasm dissipates, pilots are abandoned, and skepticism replaces optimism. The role of Finance in this moment is critical. Unlike domains driven by growth at any cost, Finance cannot afford to inflate bubbles. Its responsibility is to puncture illusions, to restore proportionality between ambition and reality, to distinguish what is genuinely transformative from what is merely fashionable, and to ensure that any adoption of agents is grounded in rigorous assessment rather than speculative enthusiasm.

4 Technology Maturity is not (... is never) the goal: the three stages of technological value

Technological maturity is often misunderstood within both academic literature and managerial practice cause it is frequently framed only as a desirable destination, a condition to be achieved, a sign that a technology has become stable, robust, predictable, and ready for widespread organizational adoption. But this framing obscures a more fundamental dynamic: maturity is not a goal but a structural turning point. It is the moment in which technological potential becomes constrained by standardization, in which the space for innovation narrows, and in which differentiation based on that technology becomes increasingly difficult. Maturity marks the transition from a source of competitive advantage to a baseline requirement, from a driver of transformation to a component of infrastructure. To understand where A(I)gents sit in this trajectory, and more importantly to understand what role Finance should expect them to play, it is essential to distinguish between three fundamentally different aims of technological value creation. These aims are not sequential stages but distinct orientations, each defining a different relationship between technology and organizational purpose. The first aim is **Efficiency (E1)**, which concerns doing existing things smarter, faster, cheaper, and more reliably. This is the domain of foundational technologies: enterprise resource planning (ERP) systems, core banking platforms, accounting ledgers, data warehouses, transaction processing engines, and deterministic automation such as robotic process automation (RPA). These technologies do not question what Finance does; they improve how Finance does it. Their primary purpose is operational stability, the reduction of errors, the acceleration of routine tasks, and the elimination of manual effort in repetitive processes. Technologies oriented toward E1 improve the existing rather than challenge it. They are essential but fundamentally conservative in nature. They make the organization more efficient at being what it already is. The second aim is **Effectiveness (E2)**, which concerns doing things differently, reinterpreting processes, introducing new analytical capabilities, and modifying information flows in ways that change not just speed but outcomes. This is the domain where intelligent automation begins to emerge, where systems move beyond mere execution to engage in interpretation, pattern recognition, anomaly detection, and decision support. Technologies in this stage are neither purely foundational nor fully transformative. They occupy a middle ground, augmenting human judgment, reshaping workflows, and enabling new forms of insight that were previously inaccessible. Intelligent automation systems, advanced analytics platforms, real-time analytics tools, and context-aware recommendation engines belong here. Their purpose is not only to accelerate existing work but to reconfigure what work means, to make visible what was hidden, to enable choices that were previously unavailable. E2 technologies question how things are done and propose alternatives. The third aim is **Entrepreneurship (E3)**, which concerns doing things for the first time, creating new patterns of action, enabling possibilities that did not exist within the previous operational paradigm. This is the frontier where agents truly reside. Technologies oriented toward E3 do not merely execute, nor do they simply optimize. They intervene in decision cycles, generate hypotheses, propose novel resolutions, interact with organizational systems in ways that traditional automation cannot anticipate, and expand the space of what is considered possible. E3 technologies are generative rather than reproductive. They do not repeat; they invent. They do not follow scripts; they improvise within constraints. Their value lies not in their ability to do known tasks better but in their capacity to discover tasks that were not previously recognized as doable. A crucial distinction emerges from this taxonomy. Foundational technologies are structurally oriented toward E1. Their success is measured by reliability, cost reduction, uptime, and consistency. They provide the substrate, the plumbing, the invisible infrastructure upon which everything else depends, but they do not, and are not designed to, generate transformational value. Intelligent automation technologies, by contrast, span E2 and E3. They are designed to reinterpret, augment, or in some cases disrupt existing cycles of work. They are tools of exploration, not just exploitation. Agents sit at the boundary of E2 and E3. They inherit the expectation of reliability characteristic of E2, the requirement that they integrate smoothly into existing systems, that they produce results that can be trusted, and that they operate within guardrails established by organizational policy. Simultaneously, they embody the generative characteristic of E3, the capacity to produce outcomes that were not explicitly programmed, to respond to novel situations with contextually appropriate actions, and to contribute to organizational intelligence in ways that exceed their original design parameters. This duality explains both their promise and their fragility. They are not merely accelerators of existing workflows, nor are they simple enhancers of human judgment operating within well-defined boundaries. They create new behavioral and informational patterns, new ways of seeing, new ways of acting. Agents do not optimize the existing; they expand the possible. And this expansion is both their value proposition and their source of risk, because expansion without governance is not progress but chaos.

5 The two evolutions that converge

The emergence of agents as a distinct category of organizational actor is not the result of a singular breakthrough, nor is it an accident of contemporary hype cycles. It represents the convergence of two long, independent technological trajectories that have developed in parallel for decades without intersecting in meaningful ways. Only recently have these trajectories matured sufficiently to meet, producing a conceptual and operational space in which new forms of semi-autonomous computational actors can exist, function, and generate value. Understanding this convergence is essential because it reveals that agents could be (in the future, for sure) not simply "more advanced AI" or "better automation." They are the synthesis of two fundamentally different evolutionary paths, each bringing its own logic, its own constraints, and its own possibilities. This synthesis is neither seamless nor inevitable; it requires deliberate architectural choices, careful governance design, and a willingness to reimagine how organizational work is structured.

5.1 The evolution of automation

The first trajectory is that of automation, which traditionally was conceived as the mechanization of human labor, the translation of manual effort into machine execution. Early automation focused on repetitive physical tasks in manufacturing environments: assembly lines, packaging systems, material handling. Over time, this conception expanded into the realm of knowledge work, giving rise to business process automation, workflow engines, and robotic process automation (RPA). The defining characteristic of traditional automation was its deterministic nature: given a specific input and a predefined set of rules, the system would produce a predictable output. This predictability was both its strength and its limitation. Automation could execute with perfect consistency, but it could not adapt, could not interpret ambiguity, could not respond to situations that fell outside its programmed parameters. As automation evolved, it moved beyond isolated tasks to encompass entire process chains, creating end-to-end orchestration of complex workflows. Systems learned to coordinate sequences of activities, manage handoffs between human and machine actors, handle exceptions through escalation mechanisms, and maintain audit trails for compliance purposes. What began as the acceleration of human activity progressively turned into the coordination of work itself. Yet even at this advanced stage, automation remained fundamentally reactive and procedural. It followed scripts, executed plans, and deferred to human judgment whenever situations became ambiguous or exceptional. This evolution did not produce autonomy in any meaningful sense, but it created the structural backbone upon which autonomy could eventually rest: an organizational environment in which computational actors were embedded in workflows, granted permissions, assigned responsibilities, and expected to perform reliably within defined parameters.

5.2 The evolution of AI

The second trajectory is that of artificial intelligence, which followed a fundamentally different path. AI did not emerge from the mechanization of labor but from the ambition to replicate or simulate cognitive functions: perception, reasoning, learning, decision-making. Its early incarnations were grounded in symbolic logic and rule-based systems, attempts to encode human knowledge explicitly in forms that machines could manipulate. These systems, known as expert systems, showed promise in narrow domains but failed to scale because the complexity of real-world knowledge exceeded what could be formally represented. The shift toward statistical learning, beginning with machine learning and accelerating with deep learning, marked a qualitative transformation. AI moved from systems that encoded knowledge to systems that inferred knowledge from data. This shift was profound because it changed the fundamental relationship between the system and the world. Instead of requiring humans to articulate everything the system needed to know, the system could discover patterns, regularities, and correlations autonomously, given sufficient data and computational resources. This discovery process was not reasoning in the classical sense but a form of empirical generalization, learning from examples rather than from principles. AI, in this statistical paradigm, was primarily a tool for interpreting existing data rather than acting upon it. Its promise resided in classification (identifying categories), prediction (forecasting outcomes), and anomaly detection (recognizing deviations from expected patterns). All of these activities were bounded by the data from which they emerged. The system could describe the world as it had been observed but could not alter it, could recognize patterns but could not generate interventions, could answer questions but could not formulate actions. The emergence of generative models, particularly large language models and diffusion-based systems, marked another qualitative transformation.

AI moved from describing the world to producing new representations of it, from following patterns to suggesting alternatives, from answering questions to formulating hypotheses, proposals, and even creative artifacts. This shift from discriminative to generative capability introduced a form of synthetic cognition, the ability to construct outputs that had never been seen before, outputs that were not simply combinations of prior examples but novel configurations informed by learned patterns. Generative AI does not reason in the sense that humans reason, nor does it understand in the sense that humans understand, but it produces outputs that exhibit properties we associate with reasoning and understanding: coherence, relevance, contextual appropriateness, and occasionally insight. This capability, however limited and brittle it may be in many contexts, is what finally allows AI to move beyond analysis toward participation in action.

5.3 The rise of the agents

Agents arise precisely at the moment in which these two trajectories converge. Automation, having evolved from task execution to process orchestration, gains the capacity for contextual judgment, informed by cognitive engines that can interpret ambiguity, recognize patterns in real time, and adapt to situations that were not explicitly anticipated in their design. AI, having evolved from pattern recognition to generative synthesis, gains the capacity for situated action, no longer confined to producing insights, recommendations, or predictions but capable of interacting directly with processes, systems, roles, and organizational workflows. This convergence creates extraordinary potential. An agent can observe an incoming stream of financial transactions, recognize patterns that suggest fraud or error, generate hypotheses about causes, consult organizational policies to determine appropriate responses, initiate corrective actions or escalations, and document its reasoning for audit purposes. This sequence of activities combines the procedural reliability of automation (the ability to execute steps consistently) with the interpretive flexibility of AI (the ability to make sense of ambiguous situations). It represents a form of organizational intelligence that is neither fully human nor purely mechanical but occupies a hybrid space between the two. However, this convergence also introduces deep structural fragility. Organizations, particularly in domains like Finance, are built around deterministic workflows, sequences of predictable steps governed by explicit rules, formal controls, periodic reconciliations, and multi-layered approval chains. The entire architecture of the Finance function is designed to eliminate ambiguity, to enforce consistency, to ensure that every action can be traced, justified, and verified. This architecture presupposes actors (whether human or computational) that execute instructions rather than interpret them, that follow rules rather than approximate them, that operate within narrow tolerances rather than adaptive ranges. Semi-autonomous actors, by contrast, introduce a fundamentally different logic. They interpret rather than execute, approximate rather than compute exactly, adapt rather than repeat, and respond to context rather than blindly follow scripts. This difference is not a minor technical detail but a profound organizational challenge. It means that the integration of agents into Finance cannot be treated as a simple technology upgrade, equivalent to migrating from one software platform to another. It requires rethinking workflows, redesigning controls, redefining roles, renegotiating responsibilities, and recalibrating the balance between efficiency and risk. The resulting tension is therefore not technological but organizational.

5.4 The challenges for Finance

Technology evolves exponentially, following Moore's Law and the dynamics of software-driven innovation, where improvements compound rapidly and capabilities that seemed impossible a few years ago become routine. Governance, however, evolves incrementally, constrained by regulatory cycles, institutional inertia, cultural norms, and the simple fact that changing how large organizations operate is slow, difficult, and risky. The function finds itself in a paradox: it is asked to adopt systems capable of judgment in structures that were designed explicitly for tools without discretion. Managing this paradox, rather than denying it or pretending it can be resolved through technical means alone, is the essential condition for the meaningful and sustainable adoption of agents in Finance. An agent, in this context, should be understood as a non-human organizational actor, one capable of participating, however partially and within defined boundaries, in cycles of observation, interpretation, and action. This does not mean that agents possess consciousness, subjective experience, genuine understanding, or intentionality in the human sense. These properties may or may not be relevant to their functionality, and debates about machine consciousness, while philosophically interesting, are largely orthogonal to the organizational question. What matters is not the "mind" of the machine but its role within a system of delegated responsibility. The distinction is conceptual, not metaphysical. An agent is defined not by its internal architecture but by its relational position within an organizational system. It is an actor to the extent that the organization treats it as a participant in a process rather than as a mere extension of human labor or a passive tool wielded by human operators. For an agent to exist as such, three conditions must converge. First, there must be an operational environment, a structured space of possible actions, data flows, process steps, and decision points, that gives shape and meaning to the agent's behavior. Without this environment, the agent is not situated and its actions have no organizational significance. Second, there must be a form of synthetic cognition, a capacity to interpret information, recognize patterns, generate responses, and adapt to variation, that enables the agent to go beyond deterministic execution. Without this capacity, the agent is merely automation, however sophisticated. Third, there must be a degree of delegated autonomy, a grant of permission to intervene in the flow of work without requiring explicit human approval for every action, that allows the agent to contribute independently to organizational outcomes. Without this autonomy, the agent is not an actor but an assistant, a tool that amplifies human action but does not participate in it. These three elements, operational environment, synthetic cognition, and delegated autonomy, do not coexist by chance. They interact, reinforce each other, and collectively create a new type of organizational reality, one in which responsibility is distributed not only among human roles but also, to a limited but significant degree, among computational actors embedded in socio-technical systems. This convergence is both an opportunity and, as emphasized repeatedly, a point of profound fragility. The opportunity lies in the potential to radically expand the capacity of Finance to process information, respond to events, detect anomalies, optimize decisions, and operate at speeds and scales that would be impossible for human teams alone. The fragility lies in the fact that organizations are structured around deterministic workflows, not adaptive collaborators. Systems evolve faster than governance. Capabilities outpace the organizational wisdom required to deploy them responsibly. The challenge is not technological but managerial, cultural, and ultimately philosophical: How do we govern actors that are not fully under our control but are also not fully autonomous? How do we assign responsibility when outcomes emerge from interactions between human and machine actors rather than from the actions of identifiable individuals? How do we audit systems that learn, adapt, and evolve rather than simply executing predefined scripts? These questions have no easy answers, but they are the questions that must be addressed if agents are to move from experimental curiosities to reliable, trustworthy, and legitimate components of the Finance function.

5.5 A philosophical note on Ontology, Epistemology, and Ethics

The introduction of agents into Finance raises questions that cannot be answered only through technical specifications, performance benchmarks, or implementation roadmaps alone. These questions become philosophical in nature, and philosophy here is not ornamental or speculative but foundational. It provides the deep grammar, the underlying structure of assumptions and principles, that determines whether delegated intelligence can be considered legitimate within a function whose authority rests entirely on trust, accountability, and adherence to norms. Three philosophical imperatives function as the structural conditions under which any form of autonomy, however limited, becomes acceptable within the Finance function. These imperatives, ontology, epistemology, and ethics, are not decorative framing devices applied after the fact to make technology sound sophisticated. They constitute the essential preconditions for agents to be integrated into organizational systems without compromising the integrity that defines the function itself. To neglect any of these imperatives is to introduce fragility, risk, and ultimately illegitimacy into the adoption process.

5.6 Ontology

What Exists for the Agent? Ontology concerns the question "What exists?" It is the domain of metaphysics, the study of being, existence, and reality. In the context of agents, ontology is not an abstract philosophical puzzle but a concrete design requirement. An agent does not navigate a neutral, objective world that simply presents itself transparently. It navigates a densely structured environment, a constructed reality made of categories, classifications, relationships, constraints, and meanings that have been defined by humans and encoded into organizational systems. For Finance, this environment includes:

accounts (assets, liabilities, equity, revenue, expenses), each defined according to accounting standards; ledgers, which structure how transactions are recorded, categorized, and aggregated; transactions, which are not mere data points but events with specific meanings, legal implications, and regulatory consequences;

entries , which link transactions to accounts according to double-entry bookkeeping principles that have existed for centuries; policies, which define thresholds, limits, approval requirements, and escalation procedures; controls, which specify what checks must be performed, what reconciliations must occur, and what documentation must be maintained;

constraints , which define what actions are permissible, what states are valid, and what transitions are allowed; exceptions, which identify situations that fall outside normal processing and require special handling; approvals, which determine who has authority to authorize specific actions; organizational roles, which define who is responsible, who must be consulted, who must be informed, and who ultimately bears accountability.

To build an agent is to define its ontology, to construct the map of the world it can perceive, interpret, and act upon. This ontology is not discovered but designed. It involves making explicit what categories matter, what relationships hold, what properties are relevant, and what distinctions are meaningful. A deficient ontology produces an agent that misclassifies reality, that sees patterns where none exist, that fails to recognize critical distinctions, and that acts on the basis of incomplete or distorted representations. In Finance, misclassification is not a theoretical problem or an inconvenience; it is a liability, a source of error that can propagate through systems, distort reports, violate regulations, and undermine trust. Consider a simple example: an agent tasked with reviewing expense reports. What does it need to "know" about expenses? It needs to understand that expenses belong to categories (travel, meals, office supplies, professional services), that each category has associated rules (per diem limits for meals, approval requirements for travel, documentation standards for professional services), that expenses are tied to cost centers and budget lines, that they must be matched against receipts and justifications, that certain patterns suggest fraud or policy violations, and that exceptions require escalation to human reviewers. Each of these elements, categories, rules, relationships, patterns, exceptions, must be explicitly represented in the agent's ontology. If any are missing, the agent's behavior becomes unreliable. Ontological design is therefore not optional. It is the first requirement for any agent that will operate in a structured domain like Finance, and the rigor of that design determines the reliability of the agent's actions.

5.7 Epistemology

How Does the Agent Know? Epistemology asks: "How does the agent know what it claims to know?" and "How justified is that knowledge?" In human terms, this is the domain of evidence, justification, reasoning, and the conditions under which belief becomes knowledge rather than mere opinion or speculation. For an agent, epistemology concerns the provenance, reliability, and justification of the information on which it bases its actions. This involves several dimensions:

1. Data provenance: where does the agent's information come from? Is it pulled directly from authoritative source systems, or is it derived, inferred, or aggregated from multiple sources? Has it been validated, or does it contain errors, inconsistencies, or missing values? Understanding provenance is essential because agents, unlike humans, cannot exercise judgment about source reliability intuitively. They must be explicitly programmed to assess it.

2. Interpretive models: what frameworks does the agent use to make sense of data? Is it applying statistical models that identify patterns in historical data? Is it using generative models that synthesize new information based on learned representations? Is it employing symbolic reasoning that follows explicit logical rules? Each of these approaches has different epistemic properties: statistical models can find correlations but cannot explain causation, generative models can produce plausible outputs that may be factually incorrect, symbolic systems can reason deductively but only within the limits of their encoded knowledge.

3. Confidence and uncertainty: what level of confidence does the agent assign to its outputs? Does it distinguish between high-certainty conclusions and tentative hypotheses? Does it communicate uncertainty to human collaborators, or does it present all outputs with equal conviction? Finance cannot tolerate systems that appear confident when they are actually guessing. Fourth, mechanisms for revision: how does the agent handle situations where its initial interpretation is challenged or contradicted by new evidence? Can it revise its conclusions, retract erroneous outputs, and learn from mistakes? Systems that cannot self-correct are systems that accumulate errors over time.

An agent without epistemology, without a coherent theory of how it knows what it knows, is a generator of noise, producing outputs that may resemble insights but have no reliable connection to reality. An agent with a weak or poorly designed epistemology is even more dangerous: it generates plausible errors, outputs that look correct, that pass surface-level tests, but that contain subtle distortions, biases, or inaccuracies that propagate downstream, influencing decisions, shaping reports, and ultimately undermining the integrity of the entire system. Finance cannot tolerate either scenario. Epistemic integrity is not optional; it is the backbone of trust, the foundation on which the legitimacy of the function rests. Every output produced by an agent must be justifiable, traceable, and auditable. This requires making epistemology explicit, designing systems that track not only what they conclude but how they concluded it, what evidence they relied on, what assumptions they made, and what level of confidence attaches to their outputs.

5.8 Ethics

What Is the Agent Allowed to Do? Ethics answers a fundamentally different question: "What is the agent allowed or not allowed to do?" Ethics, in the context of delegated intelligence, is not moral philosophy in the abstract, not a discourse about virtue, rights, or the good life. It is the explicit negotiation between what is technically possible and what is organizationally legitimate, socially acceptable, and legally permissible. It involves defining boundaries, setting limits, establishing thresholds, and specifying escalation mechanisms. In Finance, ethical design includes several components. First, operational limits of autonomy: what decisions may an agent make independently, and what decisions must always involve human judgment? There are some actions, such as approving large expenditures, authorizing significant financial commitments, or making decisions with legal implications, that should never be fully delegated to computational actors, regardless of their technical sophistication. Defining these limits is an ethical act because it reflects judgments about responsibility, accountability, and the proper distribution of authority within the organization. Second, thresholds for escalation: at what point does the agent recognize that it has encountered a situation beyond its competence and must defer to a human? This threshold may be defined in terms of financial magnitude (amounts above a certain level require human review), complexity (situations involving multiple interacting factors), ambiguity (cases where the agent's confidence falls below a specified level), or novelty (patterns that the agent has never encountered before). Without clear escalation thresholds, agents risk overstepping their competence, making decisions in domains where their epistemic reliability is unproven. Third, boundaries of acceptable risk: what level of error, what probability of failure, what magnitude of potential harm is the organization willing to tolerate in exchange for the benefits of delegation? This is fundamentally a risk-reward calculation, but it is also an ethical judgment because it involves deciding what risks can legitimately be imposed on stakeholders (employees, customers, shareholders, regulators) in pursuit of organizational efficiency or innovation. Fourth, transparency and auditability: what level of transparency is required for the agent's actions to be considered legitimate? Can auditors, regulators, and stakeholders trace how decisions were made? Can the agent's reasoning be reconstructed, examined, and evaluated? Transparency is not merely a technical requirement but an ethical one because it is the basis on which accountability becomes possible. Without transparency, responsibility dissolves; when things go wrong, there is no way to determine what failed, why it failed, or who should be held accountable. Ethics, in this sense, becomes the explicit framework through which technical capability is constrained by organizational values, legal requirements, and social norms. It is not about restricting innovation for its own sake but about ensuring that innovation serves legitimate purposes, operates within acceptable boundaries, and respects the interests of all stakeholders. In Finance, where legitimacy is paramount, ethical design is not a luxury or an afterthought; it is a precondition for adoption.

Ontology, epistemology, and ethics are not independent concerns that can be addressed separately. They are interdependent dimensions of a single challenge: how to create agents that are situated (ontology), justified (epistemology), and governed (ethics) within organizational systems. Ontology without epistemology produces an agent that sees but does not understand, that recognizes entities and relationships but cannot assess the reliability or significance of what it observes. Epistemology without ontology produces an agent that reasons but has no world to reason about, that generates conclusions without grounding them in the structured reality of the Finance domain. Ethics without either produces constraints applied to a void, rules that govern nothing because the agent has no coherent way of perceiving or knowing the world it is supposed to act within. Only the combination of all three produces an agent that can be situated in a meaningful organizational role, that generates outputs whose justification can be traced and evaluated, and that operates within boundaries that reflect not only technical possibilities but also organizational values and social responsibilities. This combination is not a theoretical ideal but a practical necessity. Any violation, whether ontological (the agent misunderstands the world), epistemological (the agent's knowledge is unjustified), or ethical (the agent exceeds its legitimate authority), transforms the agent from a resource into a risk. These imperatives are not abstractions to be discussed in academic seminars and then ignored in practice. They are design requirements, criteria that must be satisfied before any agent can be responsibly deployed in Finance. The first responsibility of the Finance function is not innovation, not efficiency, not competitive advantage. It is stewardship: the careful, prudent management of resources, the maintenance of accuracy and integrity, the protection of stakeholder interests. These philosophical principles safeguard that stewardship, ensuring that the adoption of agents does not compromise the values that define the function itself.

6 The Sci-Fi Filter: the three lenses of legitimacy

The idea of an "agent" evokes images borrowed from science fiction cinema and speculative literature: autonomous entities that operate without rules, without limits, without accountability, pursuing their own goals with little regard for human authority or organizational constraints. These narratives, while compelling as entertainment, are profoundly misleading as guides to organizational reality. In the real world of Finance, this imaginative framing must be systematically dismantled before any constructive conversation can begin. A Finance function, by its very nature, cannot tolerate actors that improvise freely, that exceed their designated remit, or that generate outcomes which cannot be traced, justified, or audited. For this reason, any form of delegated intelligence must pass through three interpretative lenses that transform the abstract concept of an "agent" from a fictional construct into an operationally legitimate actor embedded within a system of rules, roles, and responsibilities. These lenses are not arbitrary filters imposed by conservative managers resistant to innovation. They are necessary conditions derived from the fundamental nature of Finance as a function grounded in accountability, trust, and regulatory compliance. Agents that cannot pass through these lenses are not agents in any meaningful organizational sense; they are experimental prototypes, research projects, or speculative technologies that have not yet matured into forms capable of integration into real operational environments.

6.1 The first lens: Process

The first lens through which any agent must be viewed is process. A process does not merely describe a sequence of actions; it defines the structured boundaries of the world in which the agent is allowed to exist, function, and generate outcomes. It establishes what enters the system (inputs, triggers, initiating events), how information flows through stages of interpretation and transformation, what constitutes normal versus exceptional cases, how deviations are managed (exception handling, escalation procedures), what outputs are expected and under what conditions, and when human intervention is required rather than optional. Without a process, an agent is not autonomous in any productive sense but directionless, free to act without context, without purpose, and ultimately without meaning. In Finance, where the legitimacy of every action depends on its alignment with established procedures, standards, and controls, an agent without process is indistinguishable from a source of uncontrolled risk. Consider the difference between an agent that operates within a well-defined expense approval process and one that is simply told to "optimize spending." The first is bounded: it knows what documents must be checked, what thresholds apply, what policies govern different expense categories, when to approve automatically and when to escalate for human review. The second is unbounded: it may optimize according to criteria that conflict with organizational priorities, may approve expenses that violate policy, may reject legitimate claims based on misunderstood patterns, or may introduce biases that were never intended. Process provides structure, and structure provides accountability. When an agent operates within a process, its actions can be audited by examining whether it followed the defined steps, applied the correct rules, escalated appropriately, and documented its reasoning. When an agent operates outside process, accountability evaporates because there is no standard against which its behavior can be evaluated.

6.2 The second lens: Role and Autonomy

The second lens concerns role and autonomy, the degree and nature of decision-making authority granted to the agent. In fictional narratives, autonomy is often portrayed as absolute: the agent either has complete freedom or is entirely subservient, with little middle ground. In real organizations, however, autonomy is never absolute. It is always a negotiated contract, shaped by the distribution of responsibility across human and computational actors, constrained by organizational hierarchy, bounded by regulatory requirements, and subject to revision based on performance and trust. This graduated model is often articulated through the framework of Observe, Decide, Act, which represents three levels of increasing autonomy. At the first level, Observe, the agent is limited to monitoring, detecting patterns, identifying anomalies, and generating alerts or recommendations. It does not make decisions; it provides information that humans use to make decisions. This is the lowest level of autonomy and therefore the safest, the least controversial, the easiest to justify to auditors and regulators. At the second level, Decide, the agent is granted authority to make decisions within predefined thresholds and according to explicit rules. For example, it might approve expense claims below a certain amount if they satisfy all policy requirements, but escalate larger claims or ambiguous cases for human review. This level introduces real delegation but constrains it within boundaries that limit exposure. At the third level, Act, the agent not only decides but executes, taking actions that have direct operational consequences: initiating transactions, updating records, communicating with external parties, triggering downstream processes. This is the highest level of autonomy and therefore requires the most rigorous governance, the clearest accountability structures, and the most robust oversight mechanisms. The key insight here is that autonomy is not binary but graduated, and the appropriate level depends on several factors: the stakes involved (higher stakes demand more human oversight), the ambiguity of the situation (more ambiguous cases require human judgment), the agent's demonstrated reliability (agents that have proven themselves over time can be granted more autonomy), and the regulatory environment (some decisions are legally required to involve human authorization). Without this clarity about role and autonomy, the organization faces two symmetrical risks: over-delegation, where agents make decisions they should not be empowered to make, leading to errors, violations, and loss of control, and under-delegation, where agents are restricted to trivial tasks that generate little value, leading to wasted investment and missed opportunities.

6.3 The third lens: Assurance

The third lens is assurance, which addresses the fundamental question: Can this agent be explained, justified, and audited? If the answer is no, if the agent is a black box whose internal reasoning cannot be reconstructed, whose decision pathways cannot be traced, whose outputs cannot be independently verified, then it cannot be legitimately adopted in Finance, regardless of how impressive its performance might be in controlled experiments. Assurance is the moral grammar of Finance, the language through which the function communicates its integrity to stakeholders: executives who rely on financial reports, boards that oversee governance, auditors who verify compliance, regulators who enforce standards, and ultimately the public whose trust in financial institutions depends on the belief that numbers are accurate, decisions are justified, and accountability is enforceable. When an agent cannot be audited, its outputs may be useful, they may even be accurate, but they cannot be legitimate because legitimacy in Finance requires more than correctness; it requires demonstrable correctness, the ability to show not just that a conclusion is right but why it is right. Assurance requires several components.

Traceability of data: every input used by the agent must be documented, time-stamped, and linked to authoritative sources. If the agent based a decision on data from a ledger, that ledger entry must be identifiable. If it used external data (market prices, exchange rates), that data's provenance must be recorded. Without traceability, audits become impossible because there is no way to determine whether the agent was working with accurate information or corrupted data. * Interpretability of reasoning: the logical steps by which the agent moved from inputs to conclusions must be reconstructable. This does not necessarily mean that every detail of a neural network's internal representations must be human-readable, but it does mean that the overall logic must be explainable in terms that domain experts can evaluate. For example, "the agent flagged this transaction as potentially fraudulent because it deviated from the account holder's historical spending pattern in terms of amount, location, and category" is interpretable. "The agent flagged this transaction because of complex interactions among millions of parameters" is not.

Reproducibility of outcomes: given the same inputs and conditions, the agent should produce the same output. This is a basic requirement of scientific rigor, but it is also essential for accountability. If an agent's behavior is non-deterministic in ways that cannot be controlled or documented, then disputes about its actions cannot be resolved because there is no objective standard for determining what the agent should have done.

Accountability for decisions: when something goes wrong, when an error occurs, when a policy is violated, there must be a clear chain of responsibility that allows the organization to determine what failed and why. This does not mean blaming the agent (which is not a moral actor and cannot be held accountable in any meaningful sense), but it does mean identifying where in the design, deployment, or oversight of the agent the breakdown occurred, whether in data quality, model training, threshold setting, exception handling, or human review processes. Without these components, assurance is an empty promise. The organization may claim that its agents are trustworthy, but it cannot demonstrate that trustworthiness, and in Finance, claims without demonstration are worthless.

Taken together, these three lenses, process, role and autonomy, and assurance, serve as a practical filter separating technological enthusiasm from organizational reality. They distinguish what is merely possible (in a laboratory or a controlled pilot) from what is acceptable (in a production environment with real stakes, real regulations, and real accountability). They delineate the boundary between efficiency (doing things faster or cheaper) and integrity (maintaining the standards that make those efficiencies legitimate). Above all, they ensure that agents do not enter Finance as anomalies, as unstructured forces disrupting a system designed around accountability, but instead as actors integrated into a coherent architecture of rules, roles, and responsibilities. Agents that pass through these lenses are not perfect; they will still make errors, encounter edge cases, and require ongoing refinement. But they are legitimate, meaning that their presence in the organization is justified, their actions are governable, and their failures can be managed rather than catastrophic.

7 The algorithm of success and failure

The introduction of agents into the Finance function requires more than technical competence, more than enthusiasm, and more than investment. It requires a framework capable of distinguishing structural viability from superficial momentum, capable of explaining why so many technological initiatives, despite genuine innovation and substantial resources, ultimately fail to deliver on their promises. For this purpose, what can be termed the "algorithm of success and failure" provides a compact yet powerful formal model that captures a recurring organizational dynamic, a sort of an empirical pattern, distilled from the early observation about how organizations absorb (or fail to absorb) technological novelty. Its value lies not in its mathematical complexity, which is minimal, but in its explanatory power, in its ability to clarify why success is so rare and failure so common, despite good intentions, capable people, and technically sound solutions. At its core, the model is built on three variables, each ranging across the interval from zero to one, representing the degree to which an organization has achieved that particular dimension of technological evolution. The first variable, I_1 **represents Innovation**, which encompasses both the acquisition of new technologies and the organizational attitude toward generating new ideas, concepts, prototypes, and possibilities. Innovation reflects curiosity, experimentation, tolerance for ambiguity, and the willingness to explore non-linear solutions that deviate from established practice. It is the realm of labs, pilot projects, proofs of concept, and visionary thinking. Innovation is essential because without it, organizations stagnate, trapped in legacy systems and outdated paradigms. But innovation alone is never sufficient. Ideas, no matter how brilliant, have no impact if they remain ideas. The second variable, I_2 **represents Implementation**, which measures the capacity to operationalize innovation through projects, deployments, system integrations, and concrete execution. Implementation is the domain of engineering, project management, technical expertise, and delivery discipline. It asks whether what was imagined can actually be built, whether prototypes can scale to production systems, whether theoretical benefits can be realized in practice.

Implementation transforms potential into artifacts, concepts into code, possibilities into functionalities. But even successful implementation does not guarantee impact, because a system that works technically may still fail organizationally if it is not absorbed into the day-to-day work of the people who are supposed to use it. The third variable, I_3 **represents Integration**, which measures the ability to embed what has been implemented into the ongoing operations of the organization: into processes, roles, controls, behaviors, incentives, performance metrics, and cultural norms. Integration is the invisible force that turns novelty into stability, experiments into standards, pilots into practices. It involves redesigning workflows to accommodate new capabilities, training people to use new tools effectively, aligning incentives so that adoption is rewarded rather than penalized, modifying governance structures to reflect new distributions of responsibility, and ensuring that the new system becomes "the way we work" rather than "that thing IT built that nobody uses." Integration is the most difficult, the most underestimated, and the most commonly neglected of the three variables. It is also the variable that determines whether technological investment generates value or waste. The formal condition for success emerges only when all three variables simultaneously reach their maximum expression, when the organization achieves full innovation (exploring and acquiring the right technologies), full implementation (building them correctly and making them operational), and full integration (embedding them into the fabric of organizational work). In symbolic form we can say that success occurs if and only if

$$\prod_{k=1}^{3} I_k = 1$$

This is a strict conjunction, a logical AND, meaning that any deficiency in any variable produces failure, or at best, partial success that falls short of the transformative potential that justified the investment in the first place. There is no compensation between variables; excellence in innovation and implementation cannot make up for deficiency in integration. The model is unforgiving in this sense, and its unforgiving nature reflects the empirical reality that most technological initiatives fail not because the technology was inadequate but because the organization could not or would not integrate it. Any deviation from the ideal configuration of (1,1,1) produces one of several predictable failure modes, patterns that recur so consistently across different contexts that they can be named, recognized, and in principle avoided, though in practice they are rarely avoided because the organizational forces that produce them are deeply entrenched.

The first failure mode occurs when

$$I_1 = 1, I_2 < 1, I_3 = 0$$

a configuration that can be called "pilot purgatory." The organization acquires technology enthusiastically, explores its possibilities, experiments with prototypes, and generates excitement about potential applications. But it never commits to full implementation. Pilots remain pilots, indefinitely extended, perpetually "under evaluation," never scaled beyond controlled environments. This pattern is common in risk-averse organizations that want to appear innovative without actually changing anything, that want to learn without committing, that want options without obligations. Over time, enthusiasm fades, budgets are redirected, champions leave, and the pilots are quietly abandoned, leaving behind nothing but PowerPoint presentations documenting what might have been.

The second failure mode occurs when

$$I_1 = 1, I_2 = 1, I_3 = 0$$

which can be called "Faraonic cost with no value." Here, the organization not only explores but also delivers. Projects are completed, sometimes with significant investment, sometimes with great technical sophistication, sometimes celebrated in press releases and executive presentations. But the delivered systems are never absorbed into real workflows. They remain peripheral, ornamental, or entirely unused despite their technical success. Users find workarounds to avoid them, managers ignore their outputs, processes are not redesigned to leverage their capabilities. The systems exist, they function, they even work as designed, but they do not matter because the organization has not integrated them into its operational reality. This is perhaps the most frustrating failure mode because so much effort, so much expense, so much expertise was invested, yet the result is waste.

The third failure mode occurs when

$$I_1 = 1, I_2 = 1, 0 < I_3 < 1$$

a configuration that represents "unrealized promises." Solutions exist, they have been deployed, and they are even partially adopted. Some people use them, some processes incorporate them, some value is generated. But fragmentation, resistance, inconsistent usage, lack of alignment, and incomplete governance prevent them from generating the expected impact. The technology works; the organization only partially cooperates. The result is underperformance: benefits that are smaller than anticipated, ROI that is lower than projected, and a lingering sense that the initiative "didn't live up to the hype," even though the fundamental problem was not the technology but the incomplete integration.

The elegance and power of the algorithm lie in its explanatory clarity. It reveals that the real bottleneck in technological evolution, and especially in the adoption of agents, is almost never innovation or implementation. These dimensions are visible, measurable, and often well-funded. Organizations know how to acquire technology; there is an entire industry (vendors, consultants, systems integrators) dedicated to helping them do so. Organizations know how to implement technology; there are established methodologies (Agile, DevOps, project management frameworks) that guide execution. But organizations, unfortunately, systematically underestimate what integration requires. Integration is invisible work: it does not produce tangible artifacts, it does not follow clear milestones, it cannot be outsourced to vendors or delegated to IT departments. It requires changing how people work, which means overcoming inertia, addressing resistance, negotiating with stakeholders who benefit from the status quo, aligning incentives that were designed for a different operational model, and sustaining change over months or years rather than weeks. For agents specifically, whose value emerges only when they become stable, reliable actors within organizational systems rather than experimental tools used occasionally by enthusiasts, integration is not an optional final step that can be deferred or deprioritized. It is the condition that transforms a technological experiment into an organizational transformation. Agents that are not integrated remain curiosities, demonstrations of what is possible but not evidence of what is actual. The algorithm of success and failure is therefore also an algorithm of maturity. It forces Finance leaders to confront the uncomfortable truth that adoption is not a sequence of technology deployments but a sequence of organizational assimilations, and assimilation is slow, difficult, and often painful. Technology can be acquired quickly, licenses purchased, systems installed, training sessions conducted. Integration must be earned through structural change, through the patient, persistent work of redesigning processes, recalibrating controls, renegotiating roles, and reshaping the habits, norms, and expectations that constitute organizational culture.

8 From I $\rightarrow$ T $\rightarrow$ π: the impact on people

Technological evolution is always, simultaneously, an anthropological evolution. The way individuals and teams understand, adopt, reinterpret, and resist new tools determines the real magnitude of transformation far more than the capabilities of the tools themselves. Technologies do not act on organizations from the outside; they are absorbed, adapted, and transformed by the people who use them, and in turn, those people are changed by the technologies they adopt. In the case of agents, this dynamic is even more pronounced because agents do not simply replace human work, do not merely automate tasks that humans used to perform manually. They transform the structure of work, its cognitive composition, its temporal rhythm, and the nature of human expertise itself. The evolution of competencies required to function effectively in an agent-augmented environment can be described through three archetypal shapes, each representing a different model of professional development and a different relationship between specialized expertise and broader capabilities.

8.1 The I-shaped profile

represents the traditional professional built on deep, domain-specific expertise that extends vertically into a single discipline without significant breadth across other domains. In Finance, this is the expert in cost accounting, or the specialist in treasury management, or the controller who knows every detail of financial close processes but has limited knowledge of strategic planning, data analytics, or technology architecture. The I-shape is a vertical, focused form of knowledge: robust within its domain, capable of handling complexity and ambiguity within that domain, but also exposed to the risk of isolation, the risk of becoming disconnected from adjacent functions, unable to collaborate effectively across boundaries, and vulnerable to obsolescence if the domain itself is disrupted by technological or regulatory change. For decades, the I-shape was the dominant paradigm in Finance, and for good reason. Specialized expertise was essential in a world where work was divided into distinct functions, where interfaces between functions were managed through formal handoffs, and where each professional could focus narrowly on their area of responsibility without needing to understand the broader system. But as organizations became more interconnected, as processes began spanning multiple functions, and as the speed of change accelerated, the limitations of the I-shape became apparent. Specialists who could not communicate effectively with non-specialists, who could not understand the broader context in which their work was situated, and who could not adapt to changes outside their narrow domain became bottlenecks rather than assets.

8.2 The T-Shape

In response to these limitations, the T-shaped professional emerged as a more adaptive model. Here, the vertical axis of deep expertise, the "stem" of the T, is complemented by a horizontal axis of transversal competencies, the "crossbar" of the T, representing breadth across multiple domains. This horizontal dimension includes communication skills (the ability to explain complex technical concepts to non-experts), collaboration skills (the ability to work effectively in cross-functional teams), organizational fluency (understanding how different parts of the organization work and how decisions are made), contextual interpretation (recognizing how one's own work fits into broader strategic objectives), and a degree of literacy in adjacent domains (enough knowledge of IT, operations, strategy, or marketing to have informed conversations with specialists in those areas). The T-shaped professional can navigate complexity not by mastering every domain but by understanding enough about each to collaborate effectively, to ask the right questions, to recognize when expertise from another domain is needed, and to integrate insights from multiple perspectives. This profile became increasingly valuable as organizations moved toward matrix structures, cross-functional projects, and agile methodologies that required continuous coordination across boundaries that I-shaped specialists struggled to cross. But even the T-shape, while more adaptive than the I-shape, is becoming insufficient in an environment where intelligent systems, including agents, are embedded throughout organizational processes. The reason is simple: the T-shape assumes that technology is a tool wielded by humans, a means to an end, but not an active participant in work itself. It does not prepare professionals to engage with computational actors that observe, interpret, decide, and act within the same processes where humans operate.

8.3 The π-Shape or the augmented professional

, with the advent of digital transformation, data analytics, machine learning, and now agent systems, a third shape becomes necessary: the $\pi - shaped$ professional. The name derives from the Greek letter π, which visually consists of two vertical bars connected by a horizontal top, suggesting a structure that preserves the depth of the vertical axis (I), the breadth of the horizontal axis (T), but adds a second vertical dimension representing technological and analytical literacy. This does not mean that every Finance professional must become a software developer, a data scientist, or an AI engineer. That expectation would be unrealistic and counterproductive. What it does mean is that professionals must develop a level of understanding sufficient to engage with intelligent systems as cognitive partners rather than as opaque black boxes. This includes understanding the logic of how these systems work: what they can and cannot do, what kinds of reasoning they perform, what kinds of inputs they require, what kinds of outputs they produce, and what kinds of errors they are prone to making. It includes recognizing their limits and biases: that machine learning models can perpetuate historical patterns including discriminatory ones, that generative models can produce plausible but factually incorrect outputs, that optimization algorithms can find solutions that satisfy narrow objectives while violating broader constraints. It includes interpreting their outputs critically: not accepting recommendations blindly but evaluating them in context, checking them against domain knowledge, and recognizing when they make sense and when they do not. It includes evaluating their epistemic reliability: understanding the difference between a model trained on millions of examples and one trained on hundreds, between a system that reports confidence levels and one that does not, between an algorithm whose logic is transparent and one that operates as a black box. Most importantly, the π-shaped professional understands how to collaborate with agents, how to delegate appropriately, when to trust and when to override, how to provide feedback that improves system performance, and how to maintain accountability in an environment where decisions emerge from interactions between human and computational actors rather than from the actions of identifiable individuals alone.

8.4 The epistemic value of the shift

The shift from T to π is not primarily technical; it is epistemic, meaning that it changes how professionals construct meaning, how they validate knowledge, how they make decisions, and how they assume responsibility. In a world without agents, Finance professionals relied on their own judgment, informed by data, models, and tools, but ultimately grounded in human expertise and human accountability. In a world with agents, judgment becomes distributed, shared between human and computational actors, each contributing different capabilities, each compensating for the limitations of the other. This distribution does not diminish human relevance; it redefines it. The role of the professional shifts from executor to interpreter, from operator to orchestrator, from decision-maker to decision-steward.

Humans become responsible not for making every decision themselves but for ensuring that the systems within which decisions are made operate correctly, that agents are properly designed and governed, that outputs are validated, and that accountability is maintained even when specific actions are performed by non-human actors. The essential question becomes: What does a Finance professional become when an agent becomes part of their cognitive environment? The answer is not that they become redundant, nor that they become passive overseers watching machines do all the work. They become something more complex: hybrid actors operating at the intersection of human judgment and computational capability, leveraging the strengths of both, compensating for the weaknesses of both, and ultimately responsible for ensuring that the collaboration produces outcomes that are not only efficient but also legitimate, accountable, and aligned with organizational values.

9 Returning to planet Earth

After navigating definitions, conceptual triads, philosophical imperatives, technological trajectories, governance frameworks, formal models, and anthropological transformations, a return to grounded reality is necessary. Technologies, particularly those at the frontier of current capability, invite narratives that oscillate between two extremes, neither of which is helpful for organizations attempting to make responsible decisions. On one end is visionary enthusiasm, the belief that agents represent a discontinuous leap forward, a fundamental breakthrough that will revolutionize Finance, eliminating errors, accelerating decisions, and unlocking insights that were previously inaccessible. On the other end is dystopian anxiety, the fear that agents will displace human workers, introduce uncontrollable risks, propagate biases at scale, and erode accountability to the point where organizations can no longer be held responsible for their actions because "the algorithm did it." Neither extreme is grounded in reality. Agents are not magical entities capable of dissolving complexity, nor are they autonomous *super-intelligences* poised to replace human judgment across all domains. They are not prophecies waiting to unfold, not inevitable stages in some predetermined technological arc. They are engineered artifacts, systems designed by humans, trained on data collected by humans, deployed in organizations structured by humans, and governed by rules established by humans. They possess constrained forms of cognition, limited by the data they were trained on, shaped by the architectures chosen by their designers, and bounded by the environments in which they operate. Their value emerges not from their intrinsic capabilities, impressive as those may be in certain contexts, but from how they are placed within well-designed organizational architectures that leverage their strengths, compensate for their weaknesses, and integrate them into workflows in ways that generate value without compromising integrity.

When adopted without structure, agents introduce fragility. They create single points of failure, they propagate errors at scale, they generate outputs that appear authoritative but may be unreliable, they obscure accountability, and they introduce dependencies on systems that are difficult to audit, difficult to explain, and difficult to correct when they fail. When adopted with discipline, however, agents could generate genuine capability. They enable Finance to process information at speeds and scales that were previously impossible, to detect patterns that human analysts would miss, to respond to events in real time rather than after the fact, and to allocate human attention toward high-value judgment while delegating routine interpretation and coordination to computational actors. The difference between these outcomes is not primarily a function of the technology itself but of the organizational conditions under which that technology is deployed. This is the central insight that must guide adoption decisions: technology is never sufficient on its own. Its impact depends on governance, on process design, on role clarity, on assurance mechanisms, on training, on culture, and on the sustained commitment to integration that the algorithm of success and failure identifies as the true bottleneck in technological evolution. The integration of agents into Finance requires several elements, none of which is optional, all of which are prerequisites for legitimacy. First, redesigned processes that explicitly define where agents fit, what they are responsible for, what thresholds govern their autonomy, and when human intervention is required. Second, clarified roles that distribute responsibility between human and computational actors in ways that are transparent, auditable, and aligned with organizational values. Third, recalibrated controls that ensure agents operate within acceptable boundaries, that their outputs are validated, that their errors are detected and corrected, and that their performance is monitored continuously rather than assessed once at deployment and then assumed to remain stable. Fourth, defined epistemic boundaries that specify what agents are competent to handle and what falls outside their reliable range, ensuring that overreach is prevented and that delegation is matched to capability. Fifth, a renewed articulation of human responsibility that recognizes that even when agents perform specific actions, humans remain accountable for the design, governance, and oversight of the systems within which those actions occur. Responsibility cannot be delegated to machines because machines are not moral agents, are not capable of bearing responsibility in any meaningful sense. Humans remain responsible, always and irreducibly, for what happens within organizations, even when specific tasks are performed by computational actors. Finance does not need to predict the future of AI, does not need to know whether agents will become more capable, more reliable, and more autonomous over the next decade, or whether they will reach fundamental limits beyond which further progress is incremental rather than transformative. What Finance needs to do is prepare the ground, to establish the intellectual, organizational, and governance foundations that ensure that whatever future unfolds, it unfolds in ways that can be managed responsibly. This is the fundamental purpose of the conceptual work developed throughout this paper: not to advocate for agents, not to resist them, but to prepare the intellectual and organizational infrastructure so that their adoption, when it occurs, is not an act of faith, not a gamble on untested promises, but an act of governance, a deliberate, disciplined, and accountable process of integrating new capabilities into a function whose legitimacy depends on maintaining trust, accuracy, and accountability. The future of agents in Finance will be determined not by the technology alone but by the wisdom, discipline, and rigor with which Finance leaders approach their adoption. This paper has attempted to provide the conceptual tools, the analytical frameworks, and the philosophical grounding necessary for that approach to be informed, thoughtful, and ultimately successful.